ISSN 0005-1098

automatica

The Journal of IFAC the International Federation of Automatic Control

Volume 29 Number 1 January 1993

SPECIAL ISSUE ON ROBUST CONTROL

Editor H. Kwakernaak

Guest Editors J. E. Ackermann, R. F. Curtain, P. Dorato, B. A. Francis and H. Kimura

CONTENTS

Continued on outside back cover

Pergamon Press OXFORD NEW YORK SEOUL TOKYO

automatica

The IFAC Journal

By an agreement between IFAC and Pergamon Press Ltd (the official IFAC publisher) AUTOMATICA is the official Journal of IFAC the International Federation of Automatic Control

Publishing Subscription and Advertising Offices

Subscription enquiries from customers in North America should be sent to Pergamon Press Inc 660 White Plains Road Tarrytown NY 10591 5153 U S A and for the remainder of the world to Pergamon Press Ltd Headington Hill Hall Oxford OX3 0BW U K

Published six issues per annum in January March May July September and November

Application to mail at second class postage rate is pending at Rahway NJ Postmaster send address corrections to *Automatica* c o Pergamon Press Inc 660 White Plains Road Tarrytown NY 10591 5153 U S A

Subscription Rates (including postage and insurance)

Annual institutional subscription rate (1993) £510 ($US 969 00) Personal subscription rate for IFAC Affiliates (1993) £47 ($US 89) Personal subscription rate for those whose library subscribes at the regular rate is available on request Sterling prices are definitive US dollar prices are quoted for convenience only and are subject to exchange rate fluctuation Prices include postage and insurance and are subject to change without notice Subscription rates for Japan include despatch by air and prices are available on application

automatica

The Journal of IFAC the international Federation of Automatic Control

Submission of scripts
Authors should send five copies of manuscripts for publication to the appropriate Editor and one copy to the Editor-in-Chief with copy of letter to the Editor

Pergamon Press OXFORD NEW YORK SEOUL TOKYO

Aims and Scope

Automatica publishes papers on original theoretical and experimental research and development in the control of systems, involving all facets of automatic control theory and its applications. Preferably, theoretical papers should include applications; papers dealing with components and systems should include suitable theoretical background analysis and, where appropriate, consideration of economic implications.

It is intended to publish only those papers, including many based on IFAC meeting presentations, which may be regarded as new, worthwhile contributions in this field. Papers should be intelligible to the general body of control engineers, which requires that specialized techniques, terminology, and acronyms be well defined and/or referenced.

The scope of the journal is extensive. Topics covered include: the theory and design of control systems and components encompassing robust and distributed control using geometric, optimal, stochastic and nonlinear methods, game theory, and state estimation; adaptive control including robotics, neural networks, parameter estimation and system fault detection; additional topics including artificial intelligence, fuzzy and expert systems, hierarchical and man–machine systems, all parts of systems engineering which consider the reliability of components and systems; data processing; and computers for computer-aided design, manufacturing, and control of various industrial processes, space vehicles and aircraft, ships, traffic, biomedical systems, national economies, power systems, agriculture, and natural resources.

The journal also includes the following features: Special Issues on pertinent subjects in developing areas, Survey Papers, Book Reviews and Technical Communiques.

Publishing, Subscription and Advertising Offices

Subscription enquiries from customers in North America should be sent to: Pergamon Press Inc., 660 White Plains Road, Tarrytown, NY 10591-5153, U.S.A. and for the remainder of the world to: Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, U.K.

Microform Subscriptions and Back Issues

Back issues of all previously published volumes, in both hard copy and on microform, are available direct from Pergamon Press offices.

ⓢ™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

automatica

The Journal of IFAC the International Federation of Automatic Control

SPECIAL ISSUE ON ROBUST CONTROL

Editor: H. Kwakernaak

Guest Editors: J. E. Ackermann, R. F. Curtain, P. Dorato, B. A. Francis and H. Kimura

Pergamon Press OXFORD · NEW YORK · SEOUL · TOKYO

SPECIAL ISSUE ON ROBUST CONTROL

Editor: H. Kwakernaak

Guest Editors: J. E. Ackermann, R. F. Curtain, P. Dorato, B. A. Francis and H. Kimura

CONTENTS

ISSN 0005-1098

automatica

The Journal of IFAC the International Federation of Automatic Control

Volume 29 Number 2 March 1993

CONTENTS

Continued on outside back cover

Pergamon Press

OXFORD · NEW YORK · SEOUL · TOKYO

automatica

The IFAC Journal

By an agreement between IFAC and Pergamon Press Ltd (the official IFAC publisher), AUTOMATICA is the official Journal of IFAC, the International Federation of Automatic Control

---

Publishing Subscription and Advertising Offices

Subscription enquiries from customers in North America should be sent to Pergamon Press Inc 660 White Plains Road Tarrytown NY 10591 5153 USA and for the remainder of the world to Pergamon Press Ltd Headington Hill Hall Oxford OX3 0BW UK

Published six issues/annum in January March May July September and November

Application to mail at second class postage rate is pending at Rahway NJ ostmaster send address corrections to *Automatica* c/o Pergamon Press Inc 660 White Plains Road Tarrytown NY 10591 5153 USA

Subscription Rates (including postage and insurance)

Annual institutional subscription rate (1993) £510 ($US 816 00) Personal subscription rate for IFAC Affiliates (1993) £47 ($US 75) Personal subscription rate for those whose library subscribes at the regular rate is available on request Sterling prices are definitive US dollar prices are quoted for convenience only and are subject to exchange rate fluctuation Prices include postage and insurance and are subject to change without notice Subscription rates for Japan include despatch by air and prices are available on application

Automatica, Vol. 29, No. 2

Continued from outside front cover

Continued on inside back cover

PRINTED IN GREAT BRITAIN BY BPCC WHEATONS LTD, EXETER

Continued from outside back cover

INDEXED/ABSTRACTED IN *Appl Mech Rev, Curr Cont ASCA, Aqua Abstr, Cam Sci Abstr, Curr Cont CompuMath, Curr Cont /Eng Tech & Applied Sci, Comput Cont, Eng Ind, INSPEC Data, Info Sci Abstr, Math R, Oper Res Manage Sci, PASCAL-CNRS Data, Curr Cont Sci Cit Ind, Curr Cont SCISEARCH Data, SSSA/CISA/ECA/ISMEC, Zentralblatt fur Mathematik*

ISSN 0005-1098
ATCAA9 29(2) 253 586 (1993)

PUBLISHED BY

Pergamon Press OXFORD NEW YORK SEOUL TOKYO

PRINTED IN GREAT BRITAIN BY BPCC WHEATONS LTD, EXETER

Automatica, Vol. 29, No. 1, p. 1, 1993
Pergamon Press Ltd
Printed in Great Britain.

Editorial

Introduction

IN THE 1930s and 1940s, especially during the Second World War, it was a primary objective to obtain robust control of physical systems having variable characteristics and unpredictable disturbances both internal and external. Actually, the rather recently used term 'robust' was not used in that period, but it was considered in terms of sensitivity and return difference which were well defined.

Long before the age of digital computers, design techniques were mostly limited to SISO systems. Nevertheless, rather efficient, though approximate, graphical methods were developed for designing multiple feedback loops in multivariable systems including those having sampled data and even different types of nonlinearities.

However, during the 1960s and 1970s powerful digital computers became available and attention shifted from the robust benefits of feedback control to rigorous, but previously unfeasible, exact mathematical techniques for obtaining optimal control of physical systems represented by exact, invariable mathematical models where feedback control was not needed for robustness except for reducing the effects of disturbances which were usually not even considered in early studies. Application of the resulting exact control laws to actual, variable systems, especially adaptive control systems with unmodeled dynamics, were not successful. Thus the need for robust feedback techniques was resurrected, but this time to include complex, MIMO variable systems using necessary computer design methods, the derivation of which have become a primary and practically essential area of developing research.

Therefore, the *Automatica* Editorial Board and I are very grateful to the Deputy Editor-in-Chief and his Guest Editors for creating this Special Issue on Robust Control to provide some of the latest directions and a perspective of developments in robust control techniques which will be necessary to provide the potential advantages of advanced control components and systems in the future.

George S. Axelby
Editor-in-Chief

Automatica, Vol. 29, No. 1, p. 3, 1993
Pergamon Press Ltd
Printed in Great Britain.

Editorial

Special Issue on Robust Control

THE IDEA TO publish a Special Issue of *Automatica* devoted to Robust Control came up several years ago. I first talked to Peter Dorato about it during the IFAC Congress in Munich in 1987. The proposal was approved during the *Automatica* Editorial Board meeting just before the 1990 IFAC Congress at Tallinn, and the issue now appears not long before the 1993 Congress in Sydney.

I had the pleasure of working with a fine team of Guest Editors, consisting of J. Ackermann, R. F. Curtain, P. Dorato, B. A. Francis and H. Kimura. After the calls for papers late in 1990 and early in 1991, eventually 47 papers were received, several of which had been invited. The Special Issue consists of four invited papers, eight regular papers, and six brief papers. Moreover, a bibliography on robust control is included. Three papers were accepted but eventually were not published mainly because they were too long. They will appear later in regular issues of *Automatica*.

Robust control is the most interesting current research theme in the control field. Everyone knows about the difficult phase control theory went through in the late 1950s and early 1960s, when many mistakenly believed good control to be synonymous with optimal steering. Things took a better turn with the rise of linear optimal control theory in the 1960s, which at least included the idea of feedback. It lasted until the late 1970s, however, before a significant proportion of the research force of the control community returned to the central issues of control: feedback, and the robustness it can achieve.

Obviously the Special Issue does not cover the theme of robust control exhaustively. The editorial team is satisfied, though, that a representative selection of papers was assembled.

Sadly, the issue opens with a paper whose principal author is no longer alive. The work of K. Wei is commemorated in an "In Memoriam", written by his former Ph D. adviser B. Ross Barmish.

The span of the invited papers is wide but nevertheless the papers do not cover all the aspects or methodologies for robust control system analysis and design. A paper on Quantitative Feedback Theory was invited but did not make it, in common with few submitted papers on the subject.

The invited papers are all long. Barmish and Kang survey extreme point results. The paper of M. A. Dahleh and Khammash brings the ℓ_1 story. Kaminer, Khargonekar, and Rotea put mixed H_2/H_∞ control in a convex optimization framework. Packard and Doyle need approximately 50 pages to set down their work on the complex structured singular value.

The "Call for Papers" specifically requested applications papers and case studies. The paper by Chiang, Safonov, Madden and Tekawy and Garg deal with H_∞ design studies of aircraft control problems. The subject of the design study by Milanese, Fiorio, and Malan is more unusual, and concerns the robust performance of a high accuracy calibration device. Limebeer, Kasenally and Perkins present new methodology that is subsequently applied to a distillation column control problem, which was also the center of attention at a session of the 1991 IEEE Control and Decision Conference at Brighton, U.K.

The remaining regular papers and the brief papers cover a variety of theoretical aspects of robust control, ranging from robust root clustering to nonlinear uncertain systems.

The "Bibliography on Robust Control" by Dorato, Tempo and Muscato lists and classifies several hundreds of selected books and journal articles that appeared in the period 1987–1991. It supplements earlier compilations by Dorato.

Automatica's Editorial Board hopes to serve the control community at large with this and other special issues. A Special Issue on "The Challenge of Computer Science in Industrial Applications of Control", to be published jointly with the *IEEE Transactions on Automatic Control*, is being prepared by A. Benveniste and K. J. Åström, for publication during the second half of 1993. B. Wahlberg and T. Söderström are working on a Special Issue on "Signal Processing in Control", scheduled for January 1994. The Editorial Board invites proposals for other Special Issues.

Huibert Kwakernaak
Deputy Editor-in-Chief
Automatica

Automatica, Vol 29, No 1, pp 5–6, 1993
Pergamon Press Ltd
Printed in Great Britain

In Memoriam: Kehui Wei (1946–1992)

KEHUI WEI was born in Nancheng in the Chinese province of Jiangxi in 1946. As a child, he showed high academic potential skipping two years of elementary school and demonstrating aptitude and interest in Physics and Mathematics. From a young age, he aspired to become a scientist—Albert Einstein was his role model.

From 1963 to 1968, he studied Physics (specializing in microwaves) at Chengdu Institute Radio Technology. The last two years of his studies were interrupted by the Chinese Cultural Revolution and, for the next decade, he worked in the Jin-Jiang Electronics Company in Chengdu, Sichuan. During that period, he was a technician by day and a "closet scholar" by night. Colleagues at Jin-Jiang characterized Kehui as "a short thin guy with a pair of basketball shoes hanging around his neck". When recounting these difficult days, he always put a positive spin on the events. He felt that he made enormous intellectual progress during this time and was extremely proud of his role in sharing child raising responsibilites with his wife Zhuoli—he apparently took periods of leave from his job to give undivided attention to his daughter Wei-Wei. Kehui was an outstanding father.

After the cultural revolution, he was one of the first candidates to pass a very selective national examination qualifying him for study abroad. In 1980, leaving his family behind, he commenced graduate studies in Electrical Engineering at the University of Rochester where he received the Master of Science degree in 1982. Under the guidance of B. Ross Barmish, he worked towards his Ph.D. degree from 1981 to 1984.

Kehui's time in Rochester involved day-to-day interactions with contemporaries including Kris Hollot, Ian Petersen and Alberto Galimidi. He was known in the department for doing some of his best research while pacing the hallways in a cloud of smoke. Kehui was also known for his tennis matches and Florida adventures over Christmas with his friend, Deng Zhi Fang, the son of Chinese paramount leader, Chairman Deng Xiao Ping.

In 1984, he moved with his advisor to the University of Wisconsin-Madison where he completed his Ph.D. dissertation "Robust Stabilization and Pole Assignment for Linear Time-Invariant systems" in 1986. The SISO version of this work was published in a book which includes contributed papers from the 1985 MTNS; the MIMO version appeared in 1988 in *Automatica*. During the Madison period, he interacted frequently with contemporaries including Minyue Fu and Ruxiang Qian. His spirits

were enormously uplifted when his wife and daughter travelled to the U.S.A. to join him after more than four years of separation. With Zhouli's presence, this also marked the time when he improved his dietary habits—a variety of foods were substituted for his daily fast noodles. Soon thereafter, the birth of his second daughter, Linda, marked another joyous event in his life. His Ph.D. dissertation includes a note of thanks to his wife Zhuoli who "sacrificed her career and personal life to accommodate her husband's pursuit".

Throughout the entire decade of his research in the control area, Kehui continually revisited the robust stabilization problem. He could not accept the fact that a "clean" and general solution had not been given even for the problem of simultaneous stabilizability of three LTI plants with an LTI output feedback controller. For 15 months immediately following his Ph.D. he continued this line of research as a Postdoctoral Associate under the supervision of R. K. Yedavalli at the University of Toledo in Ohio. The research effort was also expanded to include unmodelled dynamics and polynomial transformation methods. In describing Kehui in jest, colleagues from Toledo recount his fascination with fax machines. Finally, he had a simple method for quick transmission of Chinese characters to all those long lost friends and relatives back home. While in Toledo, he was also known for his joviality at social gatherings and the use of a bike as his primary mode of transportation.

Over the last four years of his life, Kehui was employed at DLR (Institute for Dynamic Flight Systems) in Oberpfaffenhofen, Germany. Working with colleagues such as Georg Gruebel and Juergen Ackermann, he continued to pursue his line of research in robust control publishing more than 10 papers in major journals. He was also known at DLR for his versatility. For example, in a short period of time, he became quite involved in the problem of instantaneous torque control of a PM stepper motors with a PWM voltage source invertor—an important topic associated with the development of a new generation of robot-gripper motors.

DLR researchers described Kehui Wei as a person who was full of creative ideas, extremely hardworking (foregoing vacations) and paying little attention to his failing health. He was optimistic about his health until the very end—carrying our reviews of journal articles up to a few days before he passed away. Given his optimism, his sudden death came as a shock to his wife and two daughters.

Throughout the entire DLR period, Kehui continued to doggedly pursue a solution to the robust stabilization problem in addition to his other responsibilities. His last piece of work appears in this issue. It is a paper written in collaboration with T. Tsujino and T. Fujii. The work was completed only a couple of months before he passed away.

On 4 June 1992, we lost an excellent scientist, a fine colleague and good friend. Kehui Wei will live on through his publications the fond memories of our interaction with him and his dream to find an elegant and practical solution to the robust stabilization problem.

B. R. Barmish
J. E. Ackermann

Automatica, Vol. 29, No. 1, pp. 7–12, 1993
Printed in Great Britain

0005–1098/93 \$6.00 + 0.00

On the Connection Between Controllability and Stabilizability of Linear Systems with Structural Uncertain Parameters*

TARO TSUJINO,† TAKAO FUJII† and KEHUI WEI†‡

Robust stabilizability is equivalent to controllability invariance for a certain class of interval systems.

Key Words—Controllability; robust control; state feedback; linear systems, control systems; system theory; control theory.

Abstract—In this paper, we investigate the problem of robust stabilization of interval systems by state-feedback. It is shown under certain conditions that robust stabilizability is equivalent to controllability invariance.

1. INTRODUCTION

In recent years, the robust stabilization problem has attracted a considerable amount of interest in the field of robust control. Various kinds of necessary and sufficient conditions have been derived so far for the existence of robust stabilizing controllers. One example is a condition given in terms of the Pick matrix using interpolation theory (Kimura, 1984). Another is a type of condition given in terms of the solutions to Riccati equations both in H_∞ problem (Doyle *et al.*, 1989) and in quadratic stabilization problem (see Khargonekar *et al.* (1990) and the references therein).

For a certain class of linear interval systems, conditions for robust stabilization are given in terms of a geometric pattern with respect to the location of uncertain parameters both in the quadratic stabilization problem (Wei, 1990) and in the robust stabilization problem (Wei, 1991, 1992). These conditions are easy to check, but not necessarily easy to understand from a system theoretic point of view. In particular, the connection of these conditions with more familiar notions in the linear system theory, e.g. controllability and so on, are not so clear. Along this line Petersen (1987) defined a notion of controllability invariance in that a linear uncertain system is controllability invariant if it is controllable in usual sense for each fixed value of uncertain parameters and discussed its connection with complete quadratic stabilizability with an arbitrary degree of stability. However, he did not succeed in making a complete clarification of the connection between these two notions.

The purpose of this paper is to investigate the robust stabilization problem for interval systems along the direction in Petersen (1987). Based on the results in Wei (1990, 1991, 1992) we restrict attention to uncertain systems having a specific structure which is described in greater detail in Section 2. Along this line we show, with the help of the results in Wei (1990, 1991, 1992), that the notion of controllability invariance plays an important role in this problem. We establish that controllability invariance is necessary and sufficient for robust stabilizability, and is necessary but not sufficient for quadratic stabilizability. Thus, the main contribution of this paper is to give an intuitively appealing interpretation for the condition for robust stabilizability given in Wei (1991, 1992).

2. SYSTEMS AND DEFINITIONS

We consider a single input time-invariant interval system (or interval system for short) (A, b), which is a set of systems described by the state equation

$$\dot{x} = Ax + bu, \tag{1}$$

where $x \in \mathbf{R}^n$ is the state; $u \in \mathbf{R}$ is the control; the entries of A and b are unknown but bounded in given compact sets; i.e. $A = \{a_{ij}\}$ and $\bar{a}_{ij} \geq a_{ij} \geq \underline{a}_{ij}$; $b = \{b_i\}$ and $\bar{b}_i \geq b_i \geq \underline{b}_i$. Note that the entries of A, b vary independently. We will

* Received in final form 7 July 1992. This paper was not presented at any IFAC meeting.
† Dept. of Control Engineering and Science, Kyushu Institute of Technology, Iizuka, Fukuoka 820, Japan.
‡ DLR-Institut, für Dynamik der Flugsysteme, Oberpfaffenhofen, D-8031/Wessling, Germany.

write $a_{ij} \equiv 0$ ($b_i \equiv 0$) if $\bar{a}_{ij} = \underline{a}_{ij} = 0$ (resp. $\bar{b}_i = \underline{b}_i = 0$). The entry a_{ij} or b_i is called a sign-invariant entry if $\bar{a}_{ij} \times \underline{a}_{ij} > 0$ or $\bar{b}_i \times \underline{b}_i > 0$ and a sign-varying entry if $\bar{a}_{ij} \times \underline{a}_{ij} < 0$ or $\bar{b}_i \times \underline{b}_i < 0$. Noting Theorem 3.1 in Wei (1990) that a robustly stabilizable system must have at least the same number of sign-invariant entries in the system matrices as the system order, we restrict our attention to a class of interval systems which is called a standard system as defined below.

Definition 1. (Wei, 1991, 1992) An $n \times (n+1)$ interval matrix M is called the *associated matrix* of the interval system (A, b) if

$$M = [A \quad b]. \tag{2}$$

Furthermore an interval system (A, b) is called a *standard system* if the associated matrix $M = \{m_{ij}\}$ has the property that m_{ii+1} is a sign-invariant entry for each i, $1 \le i \le n-1$.

Definition 2. (Wei, 1991, 1992) An interval system (A, b) is said to be *stabilizable* if there exists a linear static state-feedback control law $u = kx$ with $k \in \mathbf{R}^n$ such that the characteristic polynomial of the closed-loop system

$$f(s) = \det(sI - A - bk), \tag{3}$$

is a Hurwitz invariant polynomial; i.e. all the roots of the uncertain polynomial $f(s)$ are in the strict left half of the complex plane.

Definition 3. Consider an interval system (A, b) and assume that it is a standard system. Then the interval system (A, b) is said to be *controllability invariant* if the pair (A, b) is controllable in usual sense for any fixed value of uncertain parameters, that is,

$$\operatorname{rank}[A - sI \quad b] = n, \tag{4}$$

for every $s \in \mathbf{C}$ and every $a_{ij} \in [\underline{a}_{ij}, \bar{a}_{ij}]$, $b_i \in [\underline{b}_i, \bar{b}_i]$.

Remark 1. The notion of "controllability invariance" defined above, which is essentially the same as that defined in Petersen (1987), is a natural extension of the familiar "controllability" in the linear system theory.

3 PRELIMINARY RESULTS

First we state the definition of that condition mentioned in the Introduction, which is given in terms of a geometric pattern.

Definition 4. (Wei, 1991, 1992) An $n \times (n+1)$ matrix $P = \{p_{ij}\}$ is said to be a *pattern matrix* if every entry p_{ij} of the matrix is either 0 or 1. Let Σ denote the set of all standard systems (A, b) as in Definition 1. For a given pattern matrix P, we define Σ_P as a subset of Σ determined by the following rule: A standard interval system $(A, b) \in \Sigma_P$ if $p_{ij} = 0$ implies $m_{ij} \equiv 0$ for any i, j.

According to the above definition, in order to check if an interval system $(A, b) \in \Sigma_P$ we only need to check if it is a standard system and in addition $m_{ij} \equiv 0$ when $p_{ij} = 0$.

Definition 5. (Wei, 1991, 1992) An $n \times (n+1)$ pattern matrix $P = \{p_{ij}\}$ is said to have a *generalized anti-symmetric stepwise configuration* if the following conditions hold:

(1) $p_{ii+1} = 1$ for all $i = 1, 2, \ldots, n$.
(2) If $p \ge h+2$ and $p_{hp} = 1$, then $p_{uv} = 0$ for all $u \ge v$, $u \le p-1$ and $v \le h$.
(3) $\det(P^r) \equiv p_{12}p_{23}\cdots p_{nn+1}$, where P^r is the right submatrix of P defined by

$$P^r \triangleq \begin{bmatrix} p_{12} & p_{13} & \cdots & p_{1n+1} \\ p_{22} & p_{23} & \cdots & p_{2n+1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n2} & p_{n3} & \cdots & p_{nn+1} \end{bmatrix} \tag{5}$$

The following lemma shows a necessary condition for controllability invariance and will be used later for proving one of the two key propositions, i.e. Proposition 1 (see the Appendix for its proof).

Lemma. If every interval system (A, b) in Σ_P is controllability invariant, then the following conditions hold.

(1) If $b_k \not\equiv 0$, then $a_{ij} \equiv 0 (i \ge j,\ 1 \le j \le k,\ 1 \le i \le n)$ and $a_{nk+1} \equiv 0$. (6)

(2) If $v \ge u+2$ and $a_{uv} \not\equiv 0$, then $a_{ij} \equiv 0 (i \ge j,\ 1 \le j \le u, 1 \le i \le v-1)$ and $a_{v-1u+1} \equiv 0$. (7)

The following are two key propositions for deriving one of the main results, i.e. Theorem 1.

Proposition 1. Let P be a given pattern matrix. Every interval system (A, b) in Σ_P is controllability invariant if and only if the matrix P has a generalized anti-symmetric stepwise configuration.

A proof is given in the Appendix.

Proposition 2. (Wei, 1991, 1992) Let P be a given pattern matrix. Every interval system (A, b) in Σ_P is stabilizable if and only if the matrix P has a generalized anti-symmetric stepwise configuration.

4. MAIN RESULTS

This first main result stated below is a direct consequence of Propositions 1 and 2.

Theorem 1. Every interval system (A, b) in Σ_p is stabilizable if and only if every system (A, b) in Σ_p is controllability invariant.

Remark 2. This result has the following interpretation. Suppose a standard interval system is controllability invariant, then for each fixed values of the uncertain parameters, there exists a stabilizing feedback law which may depend on the uncertain parameter. However, the theorem guarantees that the feedback law can be chosen to be independent of the uncertain parameters, in other words, it depends only on the upper and lower bounds of the uncertain parameters. Conversely, if every standard interval system in Σ_p is robustly stabilizable, then controllability invariance must hold.

In the following we introduce another notion of robust stabilizability.

Definition 6. (Hollot and Barmish, 1980; Barmish, 1985) An interval system (A, b) is said to be *quadratically stabilizable* via linear control if there exists a linear static feedback control $u = kx$, $P > 0$ and $\alpha > 0$ such that the following condition holds: for all (A, b)

$$L(x) = x^T[A'P + PA^T]x + 2x^TPbkx \leq -\alpha \, \|x\|^2,$$

where $L(x)$ is the Lyapunov derivative for the quadratic Lyapunov function $V(x) = x^TPx$ along the trajectories of the closed loop system.

For this sense of robust stabilizability we can easily obtain a corresponding result to Theorem 1, by noting that quadratic stabilizability implies robust stabilizability.

Corollary 1. Every interval system (A, b) in Σ_p is quadratically stabilizable only if every system (A, b) in Σ_p is controllability invariant.

Remark 3. We can create an example showing that the converse statement is not always true. For example, it is easy to show that the following interval system (A, b) is controllability invariant, but not quadratically stabilizable according to the main result of Wei (1990).

Example.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & a_{22} & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ 0 \\ 1 \end{bmatrix},$$

where a_{22} and b_1 are sign-varying entries.

This system is controllability invariant, but not always quadratically stabilizable for any a_{22} and b_1.

5. CONCLUSION

Theorems 1 and Corollary 1 lead to the fact that as far as a certain class of interval systems is concerned, the notion of controllability invariance defined here is necessary and sufficient for stabilizability of the interval systems and is necessary but not sufficient for quadratic stabilizability. By this fact we have connected robust stabilizability with a natural extension of the familiar notion of controllability in the linear system theory. Thus, we have found that this notion plays an important role in this robust stabilization problem. The future research is to clarify the meaning of controllability invariance defined here in the context of robust stabilization problem and also to investigate its connection with other notions of controllability, e.g. the feedback controllability as defined in Petersen (1990).

Acknowledgements—The first two authors wish to thank Prof. P. P. Khargonekar for many helpful discussions. They are also very grateful to Prof. B. R. Barmish not only for his appreciation of the main result here but also for his hearty recommendation for publication of the result in this issue of *Automatica*.

REFERENCES

Barmish, B. R. (1985). Necessary and sufficient conditions for quadratic stabilizability of an uncertain system. *J. Opt. Theory Appl.*, **46**, 399–408.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State-space solution to standard H_2 and H_∞ control problems. *IEEE Trans. Aut. Control*, **AC-34**, 831–847.

Hollot, C. V. and B. R. Barmish (1980). Optimal quadratic stabilizability of uncertain linear systems. *Proc. 18th Allerton Conf. Communic., Control Comput.*, pp. 697–706.

Khargonekar, P. P., I. R. Petersen and K. Zhou (1990). Robust stabilization of uncertain linear systems: quadratic stability and H_∞ control theory. *IEEE Trans. Aut. Control*, **AC-35**, 356–361.

Kimura, H. (1984). Robust stabilizability for a class of transfer functions. *IEEE Trans. Aut. Control*, **AC-29**, 788–793.

Petersen, I. R. (1987). Notions of stabilizability and controllability for a class of uncertain linear systems. *Int. J. Control*, **46**, 409–422.

Petersen, I. R. (1990). The matching condition and feedback controllability of uncertain linear systems. In M. A. Kaashok, J. H. van Schuppen and A. C. M. Ran (Eds), *Robust Control of Linear Systems and Nonlinear Control*, pp. 177–187. Birkhäuser, Boston.

Wei, K. (1990). Quadratic stabilizability of linear systems with structural independent time-varying uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 268–277.

Wei, K. (1991). Stabilization of linear time-invariant interval systems via constant state feedback control. DLR Technical Report R60–91.

Wei, K. (1992). Asymptotic stabilization of single-input linear systems having unknown but bounded parameters. *Proc. ACC*, pp. 1004–1008.

APPENDIX: PROOFS

Below we denote sign-invariant entries by θ_i or θ.

A. Proof of Lemma

Suppose that the following system Σ_{cl} is controllability invariant.

$$\Sigma_{cl}: A \triangleq \begin{bmatrix} 0 & \theta_1 & \cdot & \cdot & 0 \\ \cdot & & \cdot & a_{hp} & \cdot \\ a_{uv} & & & \cdot & \cdot \\ \cdot & & & \theta_{n-1} & \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} \quad b \triangleq \begin{bmatrix} 0 \\ \cdot \\ b_k \\ \cdot \\ \theta_n \end{bmatrix}, \tag{A.1}$$

Then by Definition 3 we can assume that the rank Condition 4 holds for the case where $\theta_i = 1$ $(i = 1, \ldots, n)$ and A and b contain only two sign-varying entries, b_k and a_{uv}, or a_{hp} and a_{uv}, which vary sufficiently largely. Define the matrix

$$N \triangleq [A - sI \quad b], \tag{A.2}$$

and denote N_k as the $(n-k+1) \times (n-k+1)$ lower right hand submatrix of N.

For the former part we suppose that $a_{hp} \equiv 0$, $b_k \not\equiv 0$. Then clearly by (A.1) $\operatorname{rank}[A - sI] = n - 1$ for s taking the eigenvalues of A and hence N_1 is of full rank, or, $\det N_1 \neq 0$. We also note that the characteristic polynomial of A for the system Σ_{cl} is described by

$$(-s)^{n-u+v-1}\{(-s)^{u-v+1} + a_{uv}\} = 0, \tag{A.3}$$

for $u \geq v$. Below we find conditions for the following two cases such that $\det N_1 \neq 0$ for all the roots s of the equation (A.3), or all eigenvalues of A.

Case 1: $k = 1$, meaning that b_1 is a sign-varying entry.

We furthermore separate the case depending on the location of a_{uv}.

(1) The case with $v = 1$, meaning that there exists a sign-varying entry in the first column of A. For the matrix

$$N_1 \triangleq \begin{bmatrix} 1 & & & b_1 \\ -s & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \vdots \\ & \cdot & \cdot & 0 \\ 0 & & -s & 1 \end{bmatrix}, \tag{A.4}$$

we have

$$\det N_1 = 1 + b_1(-s)^{n-1} \tag{A.5}$$

Substituting the eigenvalues of A into s in (A.5) yields

$$\det N_1 = 1, \quad (a = 0), \tag{A.6a}$$

$$\det N_1 = 1 + b_1\{(-1)^{u+1}a_{u1}\}^{n-1/u+1}(-1)^{n-1}, \quad (s = \sqrt[u]{|(-1)^{u+1}a_{u1}|}). \tag{A.6b}$$

The latter equation implies that for $a_{u1} \not\equiv 0$ and $b_1 \not\equiv 0$, there exists a_{u1} and b_1 such that $\det N_1 = 0$. Hence if $b_1 \not\equiv 0$, then $a_{u1} \equiv 0$ $(u = 1, \ldots, n)$.

(2) The case with $v > 1$, meaning that there does not exist a sign-varying entry in the first column of A. Define

$$N_1 \triangleq \begin{bmatrix} 1 & & & & b_1 \\ -s & \cdot & & 0 & 0 \\ 0 & \cdot & \cdot & & \cdot \\ \cdot & a_{uv} & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & -s & 1 \end{bmatrix}, \tag{A.7}$$

then we have

$$\det N_1 = 1 + (-1)^{n+1}b_1(-s)^{n-u+v-2}\{(-s)^{u-v+1} + a_{uv}\}, \quad (u \neq n \text{ or } v \neq 2), \tag{A.8a}$$

$$\det N_1 = 1 + (-1)^{n+1}b_1\{(-s)^{n-1} + a_{uv}\}, \quad (u = n \text{ and } v = 2). \tag{A.8b}$$

For $s = 0$, we have

$$\det N_1 = 1, \quad (u \neq n \text{ or } v \neq 2), \tag{A.9a}$$

$$\det N_1 = 1 + (-1)^{n+1}b_1 a_{uv}(v = n \text{ and } v = 2). \tag{A.9b}$$

For $s = \sqrt[u-v+1]{(-1)^{u-v+2}a_{uv}}$, we have

$$\det N_1 = 1, \quad (u \neq n \text{ or } v \neq 2), \tag{A.10a}$$

$$\det N_1 = 1, \quad (u = n \text{ and or } v = 2). \tag{A.10b}$$

Equation (A.9b) implies that if $b_1 \not\equiv 0$, then $a_{n2} \equiv 0$, where a_{n2} is the lower left hand corner of N_1. Combining (1) and (2) as above, we see that this lemma is valid for $k = 1$.

Case 2: $k > 1$, meaning that b_k is a sign-varying entry. Define

$$N_1 \triangleq \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ -s & \cdot & \cdot & & 0 & & 0 \\ 0 & \cdot & 1 & \cdot & & & 0 \\ \cdot & \cdot & \cdot & 1 & & & b_k \\ \cdot & a_{uv} & \cdot & -s & \cdot & & 0 \\ \cdot & & & \cdot & \cdot & \cdot & 0 \\ 0 & & \cdot & \cdot & 0 & -s & 1 \end{bmatrix}, \tag{A.11}$$

N_k

then we have

$$\det N_1 = \det N_k. \tag{A.12}$$

(1) When N_k does not contain a_{uv}, i.e. $v \leq k$, we have

$$\det N_k = 1 + b_k(-s)^{n-k}. \tag{A.13}$$

Since A has nonzero eigenvalues, $b_k \not\equiv 0$ implies $a_{uv} \equiv 0$ $(v = 1, \ldots, k)$ by (A.13).

(2) When N_k contains a_{uv}, i.e. $v > k$, it becomes

$$N_k \triangleq \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & b_k \\ -s & \cdot & \cdot & & 0 & \cdot \\ \cdot & \cdot & \cdot & & & 0 \\ \cdot & a_{uv} & \cdot & \cdot & & 0 \\ \cdot & \cdot & & & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & -s & 1 \end{bmatrix}. \tag{A.14}$$

Like in Case 1, we can show that $b_k \not\equiv 0$ implies that the lower left hand corner of N_k must be zero, or $a_{nk+1} \equiv 0$.

For the latter part we consider the case of $a_{hp} \not\equiv 0$ and $b_k \equiv 0$ as described below:

$$A \triangleq \begin{bmatrix} 0 & 1 & \cdot & \cdot & 0 \\ \cdot & & \cdot & a_{hp} & \cdot \\ a_{uv} & & & \cdot & \cdot \\ \cdot & & & & 1 \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \quad b \triangleq \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}. \tag{A.15}$$

First we show that this case can be reduced to the case of $a_{hp} \equiv 0$ and $b_k \not\equiv 0$ as treated in the former part. For proving this, we need the next claim, which is easy to verify.

Claim 1. *An interval system* (A, b) *is controllability invariant if and only if an interval system* (A^+, b^+) *defined below is*

controllability invariant.

$$A^{+} \triangleq \begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix}, \quad b^{+} \triangleq \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{A.16}$$

Considering the case of $u \le v$, $u \le p-1$, and define (A_{hp}, b_{hp}) as follows.

$$\Sigma_{hp} : A_{hp} \triangleq \begin{bmatrix} 0 & \theta_1 & & & \\ \cdot & & \cdot & 0 & \\ \cdot & & & \cdot & \\ \cdot & a_{uv} & & & \theta_{p-2} \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} b \triangleq \begin{bmatrix} 0 \\ a_{hp} \\ \cdot \\ 0 \\ \theta_{\Gamma} \end{bmatrix}. \tag{A.17}$$

By noting the similarity of (A.15) and (A.16), and successive use of Claim 1 we see that controllability invariance of the system (A.15) is equivalent to controllability invariance of the system (A.17), which belongs to the class of systems as treated in the former case (i.e. $a_{hp} \equiv 0$, $b_k \not\equiv 0$). We therefore conclude from the discussion for the former part that if $a_{hp} \not\equiv 0$, then $a_{uv} \equiv 0$ for all $1 \le u \le p-1$, $1 \le v \le h$ and $a_{p-1h+1} \equiv 0$. This completes the proof of Lemma.

B. *Proof of Proposition* 1

Necessity. Suppose that a standard system described below is controllability invariant

$$A \triangleq \begin{bmatrix} a_{11} & \theta_1 & a_{13} & \cdot & a_{1n} \\ a_{21} & \cdot & \theta_2 & \cdot & \cdot \\ \cdot & & \cdot & \cdot & a_{n\ 2n} \\ \cdot & & & \cdot & \theta_{n-1} \\ a_{n1} & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix}, \quad b \triangleq \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ b_{n-1} \\ \theta_n \end{bmatrix} \tag{A.18}$$

This implies the two conditions of Lemma, which include Condition 2 of Definition 5 as a subset. It thus remains to show Condition 3 of Definition 5 under the assumption, i.e.

$$\operatorname{rank}[A - sI \quad b] = n, \tag{A.19}$$

for every $s \in \mathbf{C}$ and every $a_{ij}, b_i \in \mathbf{R}$. Taking $s = 0$ in (A.19) yields

$$\operatorname{rank}[A \quad b] = n, \tag{A.20}$$

for any a_{ij}, b_i. By considering the case where the first column of A take a value of 0, we see

$$\det \begin{bmatrix} \theta_1 & a_{13} & \cdot & a_{1n} & b_1 \\ a_{22} & \theta_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & a_{n-2n} & \cdot \\ \cdot & \cdot & & \cdot & b_{n-1} \\ a_{n2} & \cdot & \cdot & \cdot & \theta_n \end{bmatrix} \triangleq \det \bar{M} \ne 0, \tag{A.21}$$

for any a_{ij}, b_i. Expanding $\bar{M}$ with respect to the first column of $\bar{M}$ yields

$$\det \bar{M} = \theta_1 \Delta(\theta_1) + a_{22} \Delta(a_{22}) + \cdots + a_{n2} \Delta(a_{n2}) \ne 0, \tag{A.22}$$

where $\Delta(*)$ is the cofactor of $*$. All terms of the above equation except for the first term contain uncertain entries, a_{ij}, varying independently, hence these terms must be equal to 0. Repeating the same discussion for the remaining first term leads to the following equation.

$$\det \bar{M} = \theta_1 \theta_2 \cdots \theta_n. \tag{A.23}$$

Therefore the conditions of Definition 5 hold for the system (A, b), namely this system has a generalized anti-symmetric stepwise configuration.

Sufficiency. Suppose that M has a generalized anti-symmetric stepwise configuration. Then by Condition 2 of Definition 5 the matrix

$$N \triangleq [A - sI \quad b], \tag{A.24}$$

has the form:

$$N = \begin{bmatrix}
-s & \theta & * & * & \cdot & \cdot & \cdot & \cdot & \cdot & * & * & * & 0 & \cdot & \cdot & 0 \\
0 & -s & \theta & * & & & & & & & & * & 0 & & & 0 \\
\cdot & & \cdot & \cdot & \cdot & & & & & & & \cdot & \cdot & & & \cdot \\
\cdot & & & \cdot & \cdot & \cdot & & & & & & \cdot & \cdot & & & \cdot \\
0 & & & & -s & \theta & * & * & \cdot & \cdot & * & * & 0 & & & 0 \\
0 & & & & 0 & *-s & \theta & 0 & 0 & 0 & 0 & 0 & 0 & & & 0 \\
\cdot & & & & \cdot & \cdot & \cdot & \theta & 0 & & & & & & & 0 \\
\cdot & & & & \cdot & \cdot & & \cdot & \cdot & \cdot & & & & & & \cdot \\
\cdot & & & & \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & & & & \cdot \\
0 & & & & 0 & * & & & & *-s & \theta & 0 & & & & 0 \\
0 & 0 & \cdot & \cdot & 0 & * & & & & & *-s & \theta & 0 & & & 0 \\
* & * & \cdot & \cdot & * & * & & & & & & *-s & \theta & \cdot & & 0 \\
* & * & \cdot & \cdot & * & * & & & & & & & *-s & \cdot & \cdot & 0 \\
* & & & & & & & & & & & & & \cdot & \cdot & 0 \\
* & * & \cdot & \cdot & * & * & \cdot & \cdot & \cdot & \cdot & * & * & * & * & *-s & \theta
\end{bmatrix} \tag{A.25}$$

where θ is a sign-invariant entry, and $*$ is a sign-varying entry.

Let the (h, p) entry of N be the lowest and rightmost sign-varying entry in the upper part of sign-invariant entries (we set $h = 0$ if there is no sign-varying entry in the upper part of sign-invariant entries). In the sequel, we prove that N has a full rank both for $s = 0$ and for $s \ne 0$, by taking Condition 3 of Definition 5 into consideration and by elementary transformation, respectively.

(1) The case with $s = 0$.

Taking $s = 0$ in (A.25) yields

$$N = \begin{array}{c@{}l}
\begin{array}{cccccccccccccccc} & & & & h & h+1 & & & & & p-1 & p & p+1 & & & \end{array} & \\
\left[\begin{array}{cccccccccccccccc}
0 & \theta & \ast & \ast & \cdot & \cdot & \cdot & \cdot & \cdot & \ast & \ast & \ast & 0 & \cdot & \cdot & 0 \\
0 & 0 & \theta & \ast & & & & & & & & \ast & 0 & & & 0 \\
\cdot & & \cdot & \cdot & \cdot & & & & & & & \cdot & \cdot & & & \cdot \\
\cdot & & & \cdot & \cdot & \cdot & & & & & & \cdot & \cdot & & & \cdot \\
0 & & & & 0 & \theta & \ast & \ast & \cdot & \cdot & \ast & \ast & 0 & & & 0 \\
0 & & & & 0 & \ast & \theta & 0 & 0 & 0 & 0 & 0 & 0 & & & 0 \\
\cdot & & & & \cdot & \cdot & \cdot & \theta & 0 & & & & & & & 0 \\
\cdot & & & & \cdot & \cdot & & \cdot & \cdot & \cdot & & & & & & \cdot \\
\cdot & & & & \cdot & \cdot & & & \cdot & \cdot & \cdot & & & & & \cdot \\
0 & & & & 0 & \ast & & & & \ast & \theta & 0 & & & & 0 \\
0 & 0 & \cdot & \cdot & 0 & \ast & & & & & \ast & \theta & 0 & & & 0 \\
\ast & \ast & \cdot & \cdot & \ast & \ast & & & & & & \ast & \theta & \cdot & & 0 \\
\ast & \ast & \cdot & \cdot & \ast & \ast & & & & & & & \ast & \cdot & \cdot & 0 \\
\ast & & & & & & & & & & & & & \cdot & \cdot & 0 \\
\ast & \ast & \cdot & \cdot & \ast & \ast & \cdot & \cdot & \cdot & \cdot & \ast & \ast & \ast & \ast & \ast & \theta
\end{array}\right] &
\begin{array}{l} \\ \\ \\ \\ h \\ h+1 \\ \\ \\ \\ \\ p-1 \\ p \\ \\ \\ \\ \end{array}
\end{array} \tag{A.26}$$

Eliminating the first column of N above yields the right matrix $\bar{M}$ of the associated matrix M, whose determinant is equal by Condition 3 of Definition 5 to the product of all diagonal elements $\theta\theta \cdots \theta$. So N has a full rank.

(2) The case with $s \neq 0$.

By shifting the first h columns of N to the right by one column and at the same time moving the $(h+1)$th column to the first column, we obtain

$$N = \begin{array}{c@{}l}
\begin{array}{cccccccccccccccc} & & & h & h+1 & & & & & & p-1 & p & p+1 & & & \end{array} & \\
\left[\begin{array}{cccccccccccccccc}
\ast & -s & \theta & \ast & \ast & \cdot & \cdot & \cdot & \cdot & & \ast & \ast & 0 & \cdot & \cdot & 0 \\
\cdot & 0 & -s & \theta & \ast & & & & & & & \ast & 0 & & & 0 \\
\cdot & \cdot & & \cdot & \cdot & \cdot & & & & & & \cdot & \cdot & & & \cdot \\
\ast & \cdot & & & \cdot & \cdot & & & & & & \cdot & \cdot & & & \cdot \\
\theta & 0 & & & & -s & \ast & \ast & \cdot & & \ast & \ast & 0 & & & 0 \\
\ast-s & 0 & & & & 0 & \theta & 0 & 0 & & 0 & 0 & 0 & & & 0 \\
\cdot & \cdot & & & & \cdot & \ast-s & \theta & 0 & & & & & & & 0 \\
\cdot & \cdot & & & & \cdot & & \cdot & \cdot & & & & & & & \cdot \\
\cdot & \cdot & & & & 0 & & & \cdot & & \cdot & & & & & \cdot \\
\ast & 0 & & & & 0 & & & & \ast-s & \theta & 0 & & & & 0 \\
\ast & 0 & 0 & \cdot & \cdot & \ast & & & & & \ast-s & \theta & 0 & & & 0 \\
\ast & \ast & \ast & \cdot & \cdot & & & & & & & \ast-s & \theta & \cdot & & 0 \\
\cdot & & & & & & & & & & & & \ast-s & \cdot & \cdot & 0 \\
\cdot & & & & & & & & & & & & & \cdot & \cdot & 0 \\
\ast & \ast & \cdot & \cdot & \ast & \ast & \cdot & \cdot & \cdot & & \ast & \ast & \ast & \ast & \ast-s & \theta
\end{array}\right] &
\begin{array}{l} \\ \\ \\ \\ h \\ h+1 \\ \\ \\ \\ \\ p-1 \\ p \\ \\ \\ \\ \end{array}
\end{array} \tag{A.27}$$

It is then easy to see that the determinant of $(n \times n)$ right submatrix is equal to a nonzero value $\theta^{n-h} \times (-s)^h$. By (1) and (2) as above, it follows

$$\operatorname{rank}[A - sI \quad b] = n. \tag{A.28}$$

for every $s \in \mathbf{C}$ and every a_{ij}, $b_i \in \mathbf{R}$. Hence (A, b) is controllability invariant. This completes the proof

Automatica, Vol. 29, No. 1, pp. 13–35, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00
Pergamon Press Ltd

A Survey of Extreme Point Results for Robustness of Control Systems*†

B. R. BARMISH‡ and H. I. KANG‡

Extreme point results for robustness.

Key Words—Robust stability; extreme points; robust performance; structured real uncertainty; polytopes of polynomials; affine linear uncertainty structures.

Abstract—This paper surveys a subset of the body of research which was sparked by Kharitonov's Theorem. The focal point is *extreme point results* for robust stability and robust performance. That is, we give conditions under which satisfaction of a performance specification is ascertained for a family of systems by only checking a finite subset of the extreme members of this family. The results which are surveyed apply mainly to systems with structured real parametric uncertainty. In addition, a number of counterexamples are given to illustrate cases for which an extreme point result does not hold. For such cases, a solution via the so-called Edge Theorem is often possible.

1. INTRODUCTION

THIS SURVEY PAPER concentrates on a subset of the major developments in robust control over the last decade—new extreme point results which facilitate robustness analysis in the presence of real parametric uncertainty. When we refer to an *extreme point result* in this paper, we mean roughly the following: It is a result which enables us to infer that some desired property of a control system is robustly satisfied by checking the satisfaction of this property on a finite subset of the "extreme" systems which may arise. For example, if the bounding set for the uncertain parameters is an ℓ-dimensional box Q and we want to have a system with damping ratio $\zeta \leq 0.707$ for all $q \in Q$, then the following question is of interest: If the damping ratio specification is satisfied for each of the 2^ℓ parameter settings associated with the vertices of Q, does it follow that this specification is also satisfied for all $q \in Q$? If so, we say that an extreme point result holds. Most of the extreme point results obtained to date apply to the *Robust Stability Problem*. However, over the last few years, a number of extreme point results have emerged for various *Robust Performance Problems* as well. The results which we survey are confined to these two problem areas. In addition, this paper includes a number of counterexamples to conjectures which one is tempted to make. That is, in many cases, a certain robustness condition holds at the extremes but fails to hold for the entire family of interest.

To put this survey in historical perspective, we begin by noting that robustness analysis with real parametric uncertainty received considerable attention in the important pioneering work of researchers such as Horowitz (1963), Siljak (1969) and Ackermann (1980). The explicit topic of this survey, however, is clearly motivated by the seminal theorem of Kharitonov (1978a). Once Kharitonov's Theorem came to light in the control community via the papers of Barmish (1983) and Bialas (1983), the pace of robustness research on real parametric uncertainty seems to have accelerated rapidly. Finally, it is also important to mention a second line of research which is also motivated by Kharitonov's Theorem but is only minimally covered here: the line of research beginning with the celebrated Edge Theorem of Bartlett *et al.* (1988). Given the fact that inclusion of a survey of edge results would probably double the length of this exposition, we opt for a sharper focus and concentrate exclusively on extreme points; for a more complete exposition covering the entire area, see the forthcoming textbook by Barmish (1993). Since this survey deals with "robustified versions" of a number of age old problems about polynomials, an important background reference

* Received 1 November 1991; received in final form 10 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor J. E. Ackermann.

† This work was supported by the National Science Foundation under Grant Number ECS-9111570.

‡ Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706, U.S.A.

to mention is the classical book by Marden (1966).

Throughout this paper, $q = (q_1, q_2, \ldots, q_\ell) \in \mathbf{R}^\ell$ represents a vector of real uncertain parameters. For example, in the well-known track-guided bus problem of Ackermann (1985), q_1 can represent a coefficient of friction between the tires of a vehicle and the surface of the road. This parameter is uncertain because it can be rainy on one day and dry on another. For robustness analysis problems involving many uncertain parameters, a major challenge is to come up with a method of testing for satisfaction of the performance specification without having to deal with a large number of contingencies. A naive approach to robustness analysis involves carrying out a classical analysis for many combinations of $(q_1, q_2, \ldots, q_\ell)$ obtained by gridding the admissible range of each q_i. In practice, however, this approach becomes computationally intractable as the number of uncertainties ℓ increases. Furthermore, if one wants to know how various controller parameters affect performance, it can be argued that this important information is lost in a gridding process. In fact, when carrying out robustness analysis, even an "intelligent" gridding process such as a branch-and-bound algorithm encounters this difficulty. The desire to avoid gridding further motivates the search for extreme point results.

By and large, the body of extreme point results which have emerged applies to feedback control problems with *structured real uncertainty*. Moreover, almost all extreme point results apply to either *independent uncertainty structures* or *affine linear uncertainty structures*. On a few occasions, *multilinear uncertainty structures* are also considered in the sequel. We now explain these terms.

Uncertainty structures

The paradigm for the discussion to follow begins with an uncertain SISO plant. Rather than using the "usual" $P(s)$ notation to describe the plant, we use the notation $P(s, q)$ to emphasize the dependence of plant coefficients on the uncertain parameter vector q; we write

$$P(s, q) = \frac{N(s, q)}{D(s, q)},$$

where $N(s, q)$ and $D(s, q)$ are *uncertain polynomials*. That is, each of these polynomials has coefficients which are functions of q. To describe an uncertain polynomial, it is convenient to write

$$p(s, q) = \sum_{i=0}^{n} a_i(q)s^i.$$

An uncertain polynomial is said to have an *affine linear uncertainty structure* if $a_i(q)$ is an affine linear function for $i = 0, 1, 2, \ldots, n$; that is, a linear function plus a constant. For example, observe that $a_i(q) = 4q_1 + 6q_2 - 7q_3 + 5$ is an affine linear function. We say that the uncertain plant $P(s, q)$ has an affine linear uncertainty structure if both the numerator $N(s, q)$ and the denominator $D(s, q)$ each have affine linear uncertainty structures.

A special case of an affine linear uncertainty structure is an *independent uncertainty structure*. In this case, each component q_i of q enters into only one coefficient. For example, the uncertain polynomial $p(s, q) = s^3 + (4 + q_1 + 6q_2)s^2 + (7 + q_3 + 2q_4)s + (5 + q_5)$ has an independent uncertainty structure. When dealing with an independent uncertainty structure, we generally lump uncertainties and simply write

$$p(s, q) = \sum_{i=0}^{n} q_i s^i.$$

Notice that in this case, we allow a zeroth component q_0 of q. In the case of rational functions, the uncertain plant $P(s, q)$ is said to have an independent uncertainty structure if both numerator $N(s, q)$ and denominator $D(s, q)$ have independent uncertainty structures and, in addition, no uncertain parameter entering into the numerator enters into the denominator and vice versa. To emphasize this uncertainty decoupling between numerator and denominator, we introduce a second uncertain parameter vector $r \in \mathbf{R}^{n+1}$ and write

$$P(s, q, r) = \frac{N(s, q)}{D(s, r)} = \frac{\sum_{i=0}^{m} q_i s^i}{\sum_{i=0}^{n} r_i s^i}.$$

We conclude this subsection with one last definition: The uncertain polynomial $p(s, q) = \sum_{i=0}^{n} a_i(q)s^i$ is said to have a *multilinear uncertainty structure* if $a_i(q)$ is a multilinear affine function for $i = 0, 1, 2, \ldots, n$; that is, if all but one component q_k of q is fixed, then $a_i(q)$ is affine linear in q_k. A simple example of a multilinear affine coefficient function is given by $a_i(q) = 5q_1q_2q_3 + 2q_2 - 6q_3 + 4q_2q_3 + 5$.

2 FURTHER MOTIVATION

Given that we restrict most of our attention in this survey to affine linear and independent uncertainty structures, it can rightfully be argued that the problems discussed here are special cases of more general problems considered elsewhere in the literature. For example, as

pointed out by Chen *et al.* (1992), a robust stability problem with affine linear uncertainty structure can be viewed as a rank one μ problem. Hence, it is correct to say that many of the results surveyed here are solutions to rather special cases of problems which can be attacked using the method of computation associated with theories such as those in Doyle (1982) or Safonov (1982). This leads us to consider two questions.

The first question which we address is: Are there advantages associated with extreme point solution when other techniques are available? To answer this question, we begin by noting that the computational complexity associated with extreme point solution can be dramatically lower than other solution methods. Perhaps the best example illustrating this point is Kharitonov's Theorem (see Section 5). For an nth order polynomial with independent uncertainty structure and interval bounds for the uncertain parameters q_i, Kharitonov's Theorem tells us that only four polynomials need to be tested to ascertain robust stability.

Without knowing Kharitonov's Theorem, one might erroneously conclude that a difficult nonlinear program must be solved in order to ascertain robust stability. To illustrate, let $\Delta_i(q)$ denote the ith leading principal minor of the Hurwitz matrix for the uncertain polynomial $p(s, q)$ and observe that $\Delta_i(q)$ is typically a nonlinear function of q. In view of the fact that robust stability is equivalent to positivity of all minors $\Delta_i(q)$, we can consider a nonlinear programming problem with objective function

$$J(q) = \min_i \Delta_i(q),$$

and constraint $q \in Q$. Clearly, robust stability is guaranteed if and only if

$$\min_{q \in Q} J(q) > 0.$$

In other words, without knowing Kharitonov's Theorem, common sense considerations might lead to an overly complicated nonlinear programming formulation. A similar motivation can be given for some of the special cases of μ theory. Once one recognizes the rank one structure, computational tractability is achieved in the sense that a "precise" global solution to the μ optimization is guaranteed via convex programming.

Further motivation for development of an extreme point theory is derived from the fact that only a few special cases of the robust stabilization problem have been solved for systems with structured real uncertainty; e.g. see the two plant solution in Vidyasagar and Viswanadham (1982), the solution for minimum phase systems in Kwakernaak (1982) and Barmish and Wei (1985), the solution for systems with a single uncertain parameter entering affinely into the plant in Khargonekar and Tannenbaum (1985) and the solution for systems satisfying a so-called matching condition as in Leitmann (1979). The key point to note is that an extreme point result reduces the problem of stabilizing infinitely many plants to the problem of stabilizing finitely many plants. If we restrict attention to compensators $C(s)$ with some fixed structure parametrized by a vector $K = (K_1, K_2, \ldots, K_p)$ and consider the Hurwitz matrix inequalities $\Delta_i(q, K) > 0$ for stability of the closed loop characteristic polynomial with q restricted to the finite set of extreme points, a synthesis becomes possible for small values of p. For example, if $C(s) = K_1 + K_2/s$ is a PI compensator, we can display the satisfaction set for these inequalities in the (K_1, K_2) plane; e.g. see Barmish *et al.* (1992) for further details.

With regard to line of research being surveyed, another important question to ask is: How rich is the class of "real world" control problems which can be attacked using a robustness theory based on affine linearity of the uncertainty structure? Although this class is admittedly "thin" in some appropriately defined problem space, we argue that in a pragmatic sense, the engineer often has little choice other than turning to existing theory for the affine linear case. To elaborate on this point, consider the following typical applications scenario: The engineer derives the characteristic polynomial $p(s, q)$ for feedback system by performing a calculation which is rather complicated (perhaps facilitated using a symbolic manipulator). The coefficient functions $a_i(q)$ involve rather "ugly" nonlinear expressions in the components q_i of q; it is not unusual for $a_i(q)$ to have hundreds of terms! Subsequently, the goal is to obtain some sort of robustness margin but it turns out that existing literature is of questionable use because the $a_i(q)$ are so complicated. In this situation, we argue that a "pragmatic" solution might involve selection of some fixed *nominal* $q^* \in Q$ followed by a robustness computation assuming independent coefficient variations: For example, one can begin with the nominal closed loop polynomial

$$p(s, q^*) = \sum_{i=0}^{n} a_i(q^*),$$

and study *percentage variations* which can be tolerated in the coefficients before instability occurs. Within the framework of this survey, this

is easily accomplished by working with the uncertain polynomial

$$p(s,q)=p(s,q^*)+\sum_{i=0}^{n} q_i s^i,$$

with variable uncertainty bounds

$$-ra_i(q^*)\le q_i\le ra_i(q^*).$$

Now, by computing the largest value of r, call it $r_{\max}$, for which robust stability is guaranteed, we obtain a *robustness margin* measured in terms of a *percentage error* with respect to *nominal coefficients*. To gain additional information about the *degree of robustness*, one can also perform a small gain calculation with the point of view that the nominal closed loop system is corrupted by unmodelled dynamics instead of complicated parametric uncertainty.

To conclude this section on motivation, it is also worth noting that affine linear uncertain structures are often quite useful for overbounding or approximating more general uncertainty structures. To illustrate, for the case of the multilinear uncertainty structures, the Mapping Theorem (for example, see Zadeh and Desoer (1963)) provides a description of the convex hull of the coefficient set. For this case, the theory to follow can be applied to this convex hull.

3 NOTATION AND TECHNICAL JARGON

In this brief section, we provide the notation and technical jargon which is used in the sequel.

Convention for real vs complex coefficients

Throughout this paper, the standing assumption is that all polynomials have real coefficients. On the few occasions when we assume complex coefficients, we state this assumption explicitly. In this regard, our point of view is as follows: in most cases, the results given here are easily modified to obtain complex coefficient analogues. Since the proof modifications for the complex coefficient case are usually straightforward, this survey avoids double citations; i.e. one citation for the first paper to provide a result and a second citation for a minor extension to the complex case.

Bounding boxes and extreme points

Throughout this paper, the uncertain parameter vectors q and r are confined to multidimensional boxes Q and R, respectively. The box description is motivated by applications in which each component of the uncertain parameter vector is known within given bounds. To describe these boxes, we assume componentwise bounds

$$q_i^-\le q_i\le q_i^+;\quad r_i^-\le r_i\le r_i^+.$$

Since the focal point of this paper is extreme point results, it is important to introduce a notation for the extreme points of Q and R. To this end we use superscript notation q^i and r^i to denote the ith extreme points of Q and R, respectively. Associated with each extreme point of the uncertainty bounding sets is a distinguished polynomial or plant. For example, if q^{i_1} is an extreme point of Q, then $p(s,q^{i_1})=\sum_{i=0}^{n} a_i(q^{i_1})s^i$ is the associated extreme polynomial. Similarly, if (q^{i_1},r^{i_2}) is an extreme point pair for $Q\times R$, then $P(s,q^{i_1},r^{i_2})$ is the associated extreme plant. If $Q\subseteq\mathbf{R}^{\ell_1}$ and $R\subseteq\mathbf{R}^{\ell_2}$, notice that there are $2^{\ell_1+\ell_2}$ extreme plants at most.

Interval polynomial families and polytopes of polynomials

If $p(s,q)$ is an uncertain polynomial having independent uncertainty structure, we henceforth refer to the set $\mathcal{P}=\{p(\cdot,q):q\in Q\}$ as an *interval polynomial family*. For convenience, we use the notation

$$p(s,q)=\sum_{i=0}^{n}[q_i^-,q_i^+]s^i.$$

When dealing with a plant $P(s,q,r)$ with independent uncertainty structure, we adopt the phraseology of Chapellat and Bhattacharyya (1989) and simply call $\mathcal{P}=\{P(\cdot,q,r):q\in Q; r\in R\}$ an *interval plant family*.

If an uncertain polynomial $p(s,q)$ has an affine linear uncertainty structure, then the *associated coefficient set*

$$a(Q)=\{(a_0(q),a_1(q),\ldots,a_n(q)):q\in Q\},$$

is a polytope in $\mathbf{R}^{n+1}$. In fact, by standard arguments from convex analysis (for example, see Rockafellar (1970)), we can express $a(Q)$ as the convex hull of the *generators* $a(q^i)$; i.e.

$$a(Q)=\operatorname{conv}\{a(q^i)\}.$$

Hence, we create an association between polytopes in $\mathbf{R}^{n+1}$ and families of polynomials with affine linear uncertainty structures. We note that every polynomial in the family $\mathcal{P}=\{p(\cdot,q):q\in Q\}$ can be expressed as a convex combination of the $p(s,q^i)$, i.e. $\mathcal{P}=\operatorname{conv}\{p(s,q^i)\}$. This further justifies calling the family $\mathcal{P}$ a *polytope of polynomials*.

Monicity, properness and invariant degree

When dealing with a robustness analysis problem, technical subtleties arise if the uncertain polynomial $p(s,q)=\sum_{i=0}^{n} a_i(q)s^i$ can drop in degree; i.e. if $a_n(q)=0$ for some admissible value of q. We opt to keep the

exposition in this survey at a more simple technical level via imposition of an invariant degree assumption wherever appropriate. To this end, when working with a feedback system, whenever needed, we assume monicity of the plant denominator and strict properness of the plant. This guarantees invariant degree of the closed loop polynomial for large classes of feedback systems with a proper compensator. For an example of a paper dealing with the subtleties of degree dropping, see Sideris and Barmish (1989).

4. INTERPLAY BETWEEN CONTROL RESULTS AND POLYNOMIAL RESULTS

Throughout the literature on extreme points, there is constantly an interplay between polynomial results and control system results—a new result involving robustness of polynomials is "dressed" in control clothing and conversely, the control clothing is often "removed" in order to isolate fundamental polynomial problems. In large measure, this interplay is motivated by a fundamental lemma which we provide below.

To motivate this lemma, let $P(s, q)$ be an uncertain plant with affine linear uncertainty structure which is connected in a feedback configuration with a compensator $C(s)$ as shown in Fig. 1. Now, if the compensator is expressed as the quotient of polynomials

$$C(s) = \frac{N_c(s)}{D_c(s)},$$

the closed loop transfer function is

$$P_{CL}(s, q) = \frac{N(s, q)D_c(s)}{N(s, q)N_c(s) + D(s, q)D_c(s)} \doteq \frac{N_{CL}(s, q)}{D_{CL}(s, q)}.$$

By taking note of the fact that the mappings from q to the coefficients of $N_{CL}(s, q)$ and $D_{CL}(s, q)$ are affine linear, the following lemma is easy to prove.

Lemma 4.1. (Affine linear uncertainty preservation.) Consider an uncertain plant connected in feedback configuration as in Fig. 1 and assume that $P(s, q)$ has an affine linear uncertainty structure. Then it follows that the uncertain closed loop transfer function $P_{CL}(s, q)$ also has an affine linear uncertainty structure.

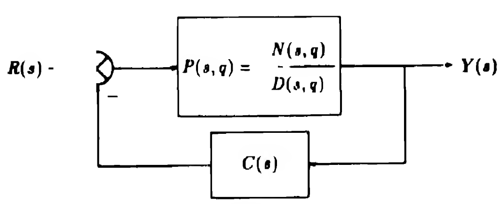


FIG. 1 Classical feedback configuration

The lemma above tells us that an affine linear uncertainty structure is preserved in going from the open loop to the closed loop. This provides significant motivation for development of robustness results at the level of uncertain polynomials with affine linear uncertainty structures; i.e. there is typically an interpretation of a polynomial result in terms of the feedback system in Fig. 1. It should also be noted that there are many variations on Lemma 4.1 which can be provided. For example, if the compensator $C(s)$ is connected in cascade with the plant and a unity feedback is used, the affine linear uncertainty structure is again preserved. More generally, the preservation of the affine linear uncertainty structure is guaranteed for all transfer functions of practical interest. For example, a simple computation of the closed loop sensitivity results in

$$S(s, q) = \frac{1}{1 + P(s, q)C(s)} = \frac{D(s, q)D_c(s)}{N(s, q)N_c(s) + D(s, q)D_c(s)}.$$

It is easy to verify that if $P(s, q)$ has an affine linear uncertainty structure, then $S(s, q)$ also has this same structure. A similar result holds for the complementary sensitivity function.

5 ROBUST STABILITY AND KHARITONOV'S THEOREM

We begin with some standard definitions: A fixed polynomial $p(s)$ is said to be *stable* if all its roots lie in the open left half plane. Given a family of polynomials $\mathcal{P} = \{p(\cdot, q) : q \in Q\}$, we say that $\mathcal{P}$ is *robustly stable* if all polynomials in $\mathcal{P}$ are stable. If all polynomials in $\mathcal{P}$ have the same degree, then we say that $\mathcal{P}$ has *invariant degree*; i.e. if $p(s, q) = \sum_{i=0}^{n} a_i(q)s^i$, then, recalling the discussion in Section 3, $\mathcal{P}$ has invariant degree if and only if $a_n(q) \neq 0$ for all $q \in Q$.

The point of departure for the literature being surveyed is Kharitonov's seminal theorem. For an interval polynomial family $\mathcal{P}$ described by $p(s, q) = \sum_{i=0}^{n} [q_i^-, q_i^+]s^i$, we extract four distinguished members

$$K_1(s) = q_0^+ + q_1^+ s + q_2^- s^2 + q_3^- s^3 + q_4^+ s$$
$$+ q_5^+ s^5 +$$
$$K_2(s) = q_0^- + q_1^- s + q_2^+ s^2 + q_3^+ s^3 + q_4^- s^4$$
$$+ q_5^- s^5 +$$

$$K_3(s) = q_0^- + q_1^+ s + q_2^+ s^2 + q_3^- s^3 + q_4^- s^4 + q_5^+ s^5 + \cdots;$$

$$K_4(s) = q_0^+ + q_1^- s + q_2^- s^2 + q_3^+ s^3 + q_4^+ s^4 + q_5^- s^5 + \cdots;$$

which are referred to as the *Kharitonov polynomials*. For example, if

$$p(s, q) = s^6 + [0.74, 1.26]s^5 + [0.75, 1.25]s^4 + [1, 2]s^3 + [2.75, 3.25]s^2 + [8.75, 9.25]s + [0.76, 10.25],$$

then the Kharitonov polynomials are

$$K_1(s) = s^6 + 1.26s^5 + 1.25s^4 + s^3 + 2.75s^2 + 9.25s + 10.25;$$

$$K_2(s) = s^6 + 0.74s^5 + 0.75s^4 + 2s^3 + 3.25s^2 + 8.75s + 0.76;$$

$$K_3(s) = s^6 + 1.26s^5 + 0.75s^4 + s^3 + 3.25s^2 + 9.25s + 0.76;$$

$$K_4(s) = s^6 + 0.74s^5 + 1.25s^4 + 2s^3 + 2.75s^2 + 8.75s + 10.25.$$

Theorem 5.1. (Kharitonov (1978a)): Let $\mathcal{P}$ be an interval polynomial family with invariant degree. Then $\mathcal{P}$ is robustly stable if and only if its associated four Kharitonov polynomials are stable.

The power of Kharitonov's theorem is derived from the fact that we can determine if $\mathcal{P}$ is robustly stable by checking only four fixed polynomials. In other words, four distinguished extreme points of Q tell the entire story. A nice proof of the theorem is given by Minnichelli *et al.* (1989); the paper by Dasgupta (1988) provides the essential geometric concepts which pave the way for the proof.

The Dasgupta geometry

Understanding of the extreme point nature of Kharitonov's solution is enhanced by considering the interval polynomial geometry of Dasgupta (1988). To this end, we temporarily freeze the frequency; say $s = j\omega_0$ with fixed $\omega_0 \in \mathbf{R}$. To avoid trivialities, we also suppose that $p(s, q^*)$ is stable for at least one $q^* \in Q$. Next, for the interval polynomial family $\mathcal{P}$, we construct the associated four Kharitonov polynomials $K_1(s)$, $K_2(s)$, $K_3(s)$ and $K_4(s)$ and generate a rectangle $p(j\omega_0, Q)$ with vertices $K_i(j\omega_0)$ as shown in Fig. 2. This rectangle is called the *value set* because it corresponds to the set of possible values that $p(j\omega_0, q)$ can assume as q ranges over Q; i.e.

$$p(j\omega_0, Q) = \{p(j\omega_0, q) : q \in Q\}.$$

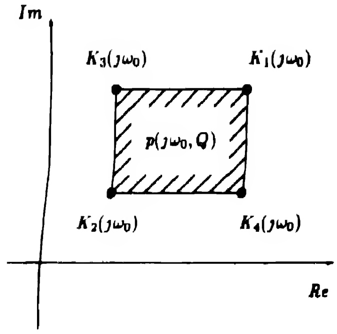


FIG. 2. Simplified description of the Kharitonov rectangle for frequency $\omega \geq 0$

Using simple ideas about continuous dependence of roots of $p(s, q)$ on q, it is easy to show that if $\mathcal{P}$ has invariant degree and at least one stable member $p(s, q^*)$, then robust stability is equivalent to satisfaction of the *Zero Exclusion Condition*

$$0 \notin p(j\omega, Q),$$

for all $\omega \geq 0$. In fact, the equivalence between zero exclusion and robust stability is not specific to interval polynomials. The Zero Exclusion Condition remains valid under much weaker conditions, e.g. if $\mathcal{P}$ has invariant degree, Q is pathwise connected and $p(s, q)$ has coefficients depending continuously on q. This idea goes back at least as far as the paper by Frazer and Duncan (1929).

The basic geometry associated with the zero exclusion condition is more fully demonstrated in Fig. 3. For $p(s, q) = [0.25, 1.25]s^3 + [2.75, 3.25]s^2 + [0.75, 1.25]s + [0.25, 1.25]$, the rectangular value set $p(j\omega, Q)$ is displayed for $0 \leq \omega \leq 1$. Using the four Kharitonov polynomials, we can also interpret zero exclusion in terms of positivity of a specially constructed frequency dependent function $H(\omega)$. With the aid of the geometry in Fig. 2, the result in Barmish (1989) is easily established: *If an interval polynomial family $\mathcal{P}$ has invariant degree and at least one stable member, then positivity of*

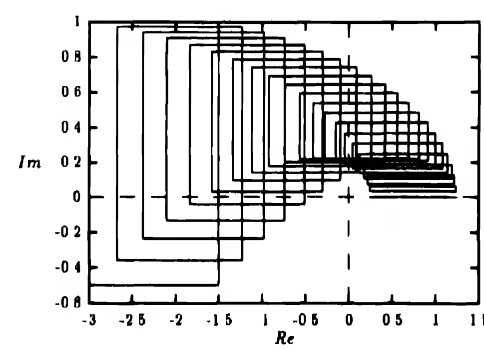


FIG. 3. Motion of the Kharitonov rectangle for $0 \leq \omega \leq 1$.

the function

$$H(\omega) \doteq \max\{\mathrm{Re}\,K_2(j\omega),\ -\mathrm{Re}\,K_1(j\omega),\ \mathrm{Im}\,K_4(j\omega),\ -\mathrm{Im}\,K_3(j\omega)\},$$

for all $\omega \geq 0$ *is equivalent to robust stability of* $\mathcal{P}$.

Polygonal value sets

The geometrical ideas above also carry over to the more general framework of polytopes of polynomials: If $\mathcal{P} = \{p(\cdot, q): q \in Q\}$ is a polytope of polynomials, then the value set $p(j\omega_0, Q)$ at frequency ω_0 is the polygon in the complex plane with generating set $\{p(\omega_0, q^i)\}$; i.e. $p(j\omega_0, Q) = \mathrm{conv}\{p(j\omega_0, q^i)\}$. The ideas underlying value set analysis go all the way back to the book by Zadeh and Desoer (1963). After laying dormant for over 20 years, we attribute the revival of these geometrical ideas to Saeki (1986) and de Gaston and Safonov (1988).

Refinements, embellishments and extensions

One of the few complex coefficient results which we mention is the followup paper by Kharitonov (1978b). For interval polynomials with complex coefficients with box bounds for both real and imaginary part, the "magic number" is eight instead of four—robust stability is guaranteed if and only if eight distinguished extreme polynomials are stable.

We now describe some results concerning the strengthening of Kharitonov's Theorem for low order polynomials. For interval polynomial families with invariant degree and order $n < 5$, special properties of the rectangular value set can be exploited to obtain the results of Anderson *et al.* (1987). If $n = 3$ and the lowest order coefficient bound satisfies $q_0 > 0$, then $\mathcal{P}$ is robustly stable if and only if the single Kharitonov polynomial $K_4(s)$ is stable. If $n = 4$ and $q_0^- > 0$, $\mathcal{P}$ is robustly stable if and only if the two Kharitonov polynomials $K_1(s)$ and $K_4(s)$ are stable. Finally, for $n = 5$, $\mathcal{P}$ is robustly stable if and only if the three Kharitonov polynomials $K_1(s)$, $K_3(s)$ and $K_4(s)$ are stable. Note that the assumption $q_0^- > 0$ above is not restrictive because a necessary condition for stability is that all coefficients have the same sign.

A simple control theoretic interpretation of Kharitonov's Theorem is given in Ghosh (1985) where an interval plant family with pure gain compensator $C(s) = K$ is considered. With configuration as in Fig. 1, it is easy to see that the denominator $D_{CL}(s, q) = KN(s, q) + D(s, q)$ of the closed loop transfer function still has an independent uncertainty structure. Hence, closed loop stability can be studied using Kharitonov's Theorem. This type of reasoning leads to an eight polynomial test for robust stability—four polynomials for $K \geq 0$ and four polynomials for $K < 0$. Furthermore, this same analysis implies that the "worst case" gain margin is achieved by a Kharitonov polynomial; see Section 9 for further discussion.

To conclude this section, we mention an extension of the theory which applies to multilinear uncertainty structures satisfying a so-called *even–odd decoupling condition*: We assume that any uncertain parameter q_k which enters into an even order coefficient $a_{2i}(q)$ of $p(s, q)$ does not enter into any odd order coefficient, and conversely, if q_k enters into an odd order coefficient $a_{2i+1}(q)$ of $p(s, q)$, then it does not enter into an even order coefficient. Using this condition, the paper by Kharitonov (1979) provides the following result: Suppose $\mathcal{P} = \{p(\cdot, q): q \in Q\}$ is a family of polynomials which has invariant degree, multilinear uncertainty structure and satisfies the even–odd decoupling condition. Then $\mathcal{P}$ is robustly stable if and only if $p(s, q^i)$ is stable for all extreme points q^i of Q. Underlying a simple proof of this result is the basic fact that the value set $p(j\omega, Q)$ remains rectangular when the even–odd decoupling condition is satisfied; see also Panier *et al.* (1989) for further extensions.

6. OTHER ROOT LOCATION REGIONS

After reading Kharitonov's Theorem, an obvious question to ask is: To what extent can extreme point results be given for root location regions other than the strict left half plane? For example, for discrete time control problems, the focal point is the unit disc. More generally, given a family of polynomials $\mathcal{P} = \{p(\cdot, q): q \in Q\}$ and a desired open root location region $\mathcal{D} \subseteq \mathbf{C}$, we say that $\mathcal{P}$ is *robustly $\mathcal{D}$-stable* if, for each $q \in Q$, $p(s, q)$ has all its roots in $\mathcal{D}$. We reserve the phrase *robust stability* for the special case when $\mathcal{D}$ is the open left half plane. For the special case when $\mathcal{D}$ is the open unit disc, we use the phrase *robust Schur stability*.

We call $\mathcal{D}$ a *weak Kharitonov region* if the following condition holds: Given any interval polynomial family $\mathcal{P}$ having invariant degree, stability of $p(s, q^i)$ for all extreme points q^i of Q implies robust $\mathcal{D}$-stability of $\mathcal{P}$. Taking n to be the degree of all polynomials in $\mathcal{P}$, we call $\mathcal{D}$ a *strong Kharitonov region* if there exists an index set $I(\mathcal{D}) \subseteq \{1, 2, 3, \ldots, 2^{n+1}\}$ with cardinality independent of n and associated labelling of extreme polynomials $\{p(s, q^i)\}$ such that $\mathcal{D}$-stability of $p(s, q^i)$ for $i \in I(\mathcal{D})$ implies robust $\mathcal{D}$-stability of $\mathcal{P}$. The understanding above is that the labelling scheme for the extreme

polynomials may depend on n. For example, in Kharitonov's Theorem, by appropriate labelling of extreme points, we can take $I(\mathscr{D}) = \{1, 2, 3, 4\}$. From the point of view of computational complexity, it is obvious that results involving strong Kharitonov regions are quite powerful when the cardinality of $I(\mathscr{D})$ is small.

When studying the robust Schur stability problem, the first obvious conjecture to make is that the open unit disc is a weak Kharitonov region. However, a number of authors provide counterexamples to such a conjecture. For example, in Bose and Zeheb (1986), the interval polynomial family described by $p(s, q) = s^4 + [-\frac{17}{8}, \frac{17}{8}]s^3 + \frac{3}{2}s^2 - \frac{1}{3}$ has all extremes with roots in the open unit disc. However, the "intermediate" polynomial $p^*(s) = s^4 + \frac{3}{2}s^2 - \frac{1}{3}$ is a member of the family and has two root $s \approx \pm j1.3025$ which are outside the unit disc. Although the open unit disc fails to be a weak Kharitonov region, the paper by Petersen (1989) indicates that there are other important classes of weak Kharitonov regions. Such classes include regions such as shifted half planes, damping cones and circles which are entirely contained in the open left half plane. Petersen also goes on to show that the intersection of any two of his regions is once again a weak Kharitonov region. In the theorem below, a much more general characterization of weak Kharitonov regions is given.

Theorem 6.1. (Rantzer (1992b)) Suppose $\mathscr{D} \subseteq \mathbf{C}$. *Then* $\mathscr{D}$ *is a weak Kharitonov region if* $\mathscr{D}$ and $1/\mathscr{D} = \{z : zd = 1 \text{ for some } d \in \mathscr{D}\}$ are convex.

In fact, the paper by Rantzer actually gives an even stronger result: *For the case of complex coefficient polynomials, convexity of both* $\mathscr{D}$ *and* $1/\mathscr{D}$ *is also necessary in order for* $\mathscr{D}$ *to be a weak Kharitonov region.* Note that Rantzer's theorem enhances our understanding of a number of facts which we already know. For example, consistent with the fact that general extreme points are not available for the robust Schur stability problem, take $\mathscr{D}$ to be the open unit disc and notice that $1/\mathscr{D} = \{z : |z| > 1\}$ is not convex. On the other hand if $\mathscr{D}$ is a half plane or a convex cone, then its reciprocals are easily seen to be convex. This is consistent with the results of Petersen (1989).

More on robust Schur stability problems

Given that the unit disc is not a weak Kharitonov region, a number of authors have considered the following question: For an interval polynomial, what reasonable strengthening of hypotheses enables us to infer robust Schur stability from Schur stability of the *extreme points? In addressing this question, the first point to note is that it is no accident that a* fourth order polynomial is used above to demonstrate that the unit disc is not a weak Kharitonov region. In fact, for monic polynomials of order $n \le 3$, the papers by Cieslik (1987) and Kraus *et al.* (1988) tell us that robust Schur stability is guaranteed if and only if all of the extreme polynomials $p(s, q^i)$ are Schur stable. For higher order polynomials, an extreme point result of Hollot and Bartlett (1986) applies to the special case when roughly half the coefficients of $p(s, q)$ are fixed. Since $p(1, q) \neq 0$ is necessary for Schur stability, without loss of generality, the assumption $\sum_{i=0} q_i^- > 0$ is imposed in the theorem below.

Theorem 6.2. (Hollot and Bartlett (1986)) Consider the discrete time interval polynomial family $\mathscr{P}$ described by $p(s, q) = \sum_{i=0}^{n} [q_i^-, q_i^+]s^i$ and in addition, assume that $\sum_{i=0}^{n} q_i^- > 0$ and $q_i^- = q_i^+$ for $i = \lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + 2, \ldots, n$ where $\lfloor n/2 \rfloor$ denotes the largest integer less than or equal to $n/2$. Then $\mathscr{P}$ is robustly Schur stable if and only if all of the extreme polynomials $p(s, q^i)$ are Schur stable.

To conclude this subsection, we mention the result of Soh (1990) for robust Schur aperiodicity: We say that the interval polynomial family $\mathscr{P} = \{p(., q) : q \in Q\}$ is *robustly aperiodic with respect to the unit disc* if, for all $q \in Q$, all roots of $p(s, q)$ are real, positive, distinct and lie in $(0, 1)$. Within this framework, Soh identifies two distinguished extreme polynomials $p(s, q^1)$ and $p(s, q^2)$ which tell the entire story; i.e. $\mathscr{P}$ is robustly aperiodic with respect to the unit disc if and only if all roots of $p(s, q^1)$ and $p(s, q^2)$ lie in $(0, 1)$ and are distinct.

Damping cones

When dealing with an nth order interval polynomial $\mathscr{P}$ and a weak Kharitonov region $\mathscr{D}$, note that a stability test can require working with as many as 2^{n+1} extreme polynomials. We now describe a pair of results which indicate that damping cones enable us to carry out a robust stability test with much smaller number of extremes. In the two theorems below, we refer to the damping cone $\mathscr{D}$ shown in Fig. 4.

Theorem 6.3. (Foo and Soh (1989)) Given a damping cone $\mathscr{D}$ and an nth order interval polynomial family $\mathscr{P}$ having invariant degree,

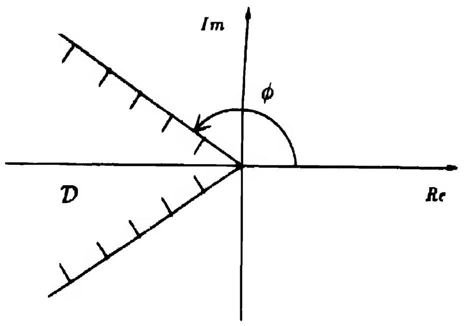


FIG 4 Damping cone as a $\mathscr{D}$ region.

there exists a distinguished subset of, at most, $2(n+1)$ extreme polynomials $\{p(s, q^1), p(s, q^2), \ldots, p(s, q^{2(n+1)})\}$ having the following property: $\mathscr{P}$ is robustly $\mathscr{D}$-stable if and only if $p(s, q^i)$ is $\mathscr{D}$-stable for $i = 1, 2, \ldots, 2(n+1)$.

The paper by Foo and Soh (1989) also includes a simple recipe for construction of the distinguished extreme polynomials. For example, for $n = 3$, there are eight distinguished extremes given by

$$\begin{aligned}
p(s, q^1) &= q_0^+ + q_1^+ s + q_2^- s^2 + q_3^+ s^3, \\
p(s, q^2) &= q_0^- + q_1^- s + q_2^+ s^2 + q_3^- s^3. \\
p(s, q^3) &= q_0 + q_1^+ s + q_2^+ s^2 + q_3^- s^3; \\
p(s, q^4) &= q_0^+ + q_1^- s + q_2^- s^2 + q_3^+ s^3; \\
p(s, q^5) &= q_0^+ + q_1^- s + q_2^+ s^2 + q_3^+ s^3; \\
p(s, q^6) &= q_0^- + q_1^+ s + q_2^- s^2 + q_3 s^3; \\
p(s, q^7) &= q_0^- + q_1^+ s + q_2^- s^2 + q_3^+ s^3; \\
p(s, q^8) &= q_0^+ + q_1^- s + q_2^+ s^2 + q_3 s^3.
\end{aligned}$$

In a later paper by Soh (1992), it is shown that only four of the eight specified complex coefficient polynomials of Soh and Berger (1988b) are required when $\mathscr{D}$ is the damping cone above.

To motivate the next theorem, we take note of the fact that the open left half plane is the only strong Kharitonov region which we have identified thus far. The next result, due to Soh and Foo (1990) and Rantzer (1990), strengthens an earlier result by Soh and Berger (1988a). We see below that there is a large class of damping cones which also enjoy the strong Kharitonov region property.

Theorem 6.4. (Soh and Foo (1990) and Rantzer (1990)) Consider an nth order interval polynomial family $\mathscr{P}$ having invariant degree and a damping cone $\mathscr{D}$ with angle $\phi = (n_1/n_2)\pi \geq \pi/2$ where n_1 and n_2 are relatively coprime positive integers. Then, there exists a distinguished subset of at most $2n_2$ extreme polynomials $\{p(s, q^1), p(s, q^2), \ldots, p(s, q^{2n_2})\}$ having the following property: $\mathscr{P}$ is robustly $\mathscr{D}$-stable if and only if $p(s, q^i)$ is $\mathscr{D}$-stable for $i = 1, 2, \ldots, 2n_2$.

It is interesting to compare Theorems 6.3 and 6.4. If the objective is to minimize the number of extremes to be tested, one cannot say that either theorem is uniformly better. For example, if $n = 5$ and $\phi = 2\pi/3$, then Theorem 6.3 requires up to 12 extremes but Theorem 6.4 requires six extremes. On the other hand, if $n = 9$ and $\phi = \frac{139}{180}\pi$, then Theorem 6.3 requires up to 20 extremes and Theorem 6.4 requires up to 360 extremes. For further results aimed at reducing the number of extremes to be tested, see Soh and Foo (1991).

7. THE EDGE THEOREM: A BRIDGE TO FURTHER RESULTS

There are a number of extreme point results in the literature whose proofs rely on the celebrated Edge Theorem of Bartlett *et al.* (1988). Loosely speaking, the Edge Theorem tells us the following: For robust $\mathscr{D}$-stability problems with affine linear uncertainty structures, we need only check $\mathscr{D}$-stability of all polynomials associated with the edges of the Q box. If all polynomials corresponding to edge points of Q are $\mathscr{D}$-stable, then the entire polynomial family is robustly $\mathscr{D}$-stable. Since the edges of Q are only one-dimensional, we obtain a dramatic reduction in computational complexity. On the downside, however, note that if $Q \subseteq \mathbf{R}^\ell$, then this box has as many as $\ell 2^{\ell-1}$ edges. We now make these ideas more precisely.

Given a polytope of polynomials $\mathscr{P} = \{p(\cdot, q): q \in Q\}$, we define its exposed edges of this set of functions via the natural isomorphism between $\mathscr{P}$ and its coefficient space representation; i.e. recalling the discussion in Section 3, if $a(q)$ is the coefficient vector for $p(s, q)$, the coefficient set $a(Q)$ is a polytope obtained by taking the convex hull of the $a(q^i)$. We now use the exposed edges of the polytope $a(Q)$ to define exposed edges of $\mathscr{P}$ in the obvious way. That is, for each a^* which lies on the exposed edge of $a(Q)$, we obtain some $q^* \in Q$ and an edge polynomial $p(s, q^*)$ in $\mathscr{P}$ with coefficient vector $a(q^*) = a^*$. When we refer to the exposed edges of $a(Q)$ above, we mean the one-dimensional exposed faces; e.g. see Rockafellar (1970). Using the affine linear uncertainty structure and some standard convex analysis arguments, it is not difficult to prove that every polynomial lying on an exposed edge of $\mathscr{P}$ is obtained from an exposed edge of Q. However, an exposed edge of Q does not necessarily induce an exposed edge of $\mathscr{P}$. We are now prepared to provide the most well-known version of the Edge Theorem.

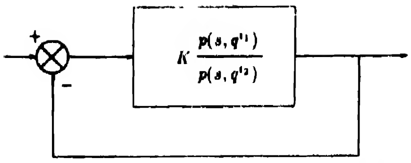


FIG. 5. Fictitious plant for solution of an edge problem.

Theorem 7.1. (Bartlett *et al.* (1988)) Let $\mathcal{P}$ be a polytope of polynomials with invariant degree and assume that $\mathcal{D} \subseteq \mathbf{C}$ is open and simply connected. Then $\mathcal{P}$ is robustly $\mathcal{D}$-stable if and only if each of the exposed edges of $\mathcal{P}$ is $\mathcal{D}$-stable.

When applying the Edge Theorem, it is often easier to work with the exposed edges of Q in lieu of the exposed edges of $\mathcal{P}$. By working with exposed edges of Q, however, we accept the possibility of redundancy in the robustness test; i.e. we may end up testing a superset of the exposed edges of $\mathcal{P}$. In summary, by working with edges, we reduce the robust $\mathcal{D}$-stability problem for a polytope of polynomials to a finite number of one-dimensional edge problems. Each of these one-dimensional edge problems can be solved by a variety of classical methods. For example, if q^{i_1} and q^{i_2} are two extreme points of Q whose convex combination describes an exposed edge of Q, then a robust $\mathcal{D}$-stability test amounts finding the roots of the polynomial

$$p_{i_1,i_2}(s,\lambda) \doteq (1-\lambda)p(s,q^{i_1}) + \lambda p(s,q^{i_2}),$$

for $\lambda \in [0,1]$. By dividing by $\lambda p(s,q^{i_2})$ above, it becomes clear that the edge problem reduces to a classical root locus plot for the *fictitious plant*

$$P_{i_1,i_2}(s) \doteq \frac{p(s,q^{i_1})}{p(s,q^{i_2})},$$

which is compensated via unity feedback as shown in Fig. 5.

8 THE 16 PLANT THEOREM

To motivate the results surveyed in this section and the next, we begin with an interval plant family $\mathcal{P}$ described by

$$P(s,q,r) = \frac{\sum_{i=0} [q_i^-, q_i^+]s^i}{\sum_{i=0}^{n} [r_i^-, r_i^+]s^i},$$

and return to the case when the desired root location region $\mathcal{D}$ is the open left half plane. Recalling the notation $P(s, q^{i_1}, r^{i_2})$ for the plant associated with the extreme point pair $(q^{i_1}, r^{i_2}) \in Q \times R$, the fundamental question which we now address is: If a compensator $C(s)$, connected as in Fig. 1, stabilizes each of the extreme plants $P(s, q^{i_1}, r^{i_2})$, does it follow that $C(s)$ robustly stabilizes the entire family $\mathcal{P}$?

Thirty two edges

In order to address this question, we bring the result of Chapellat and Bhatacharyya (1989) into play. That is, if the closed loop polynomial has invariant degree, we do not need to check all the edges in order to ascertain robust stability. It is necessary and sufficient to check a set of 32 distinguished edges. These 32 edges are dubbed *Kharitonov segments* in Chapellat and Bhattacharyya (1989); note that the number 32 is independent of m and n. These distinguished edges are obtained as follows: We freeze the plant numerator at some Kharitonov polynomial $N_{i_\bullet}(s)$ and consider four denominator edges of the form $(1-\lambda)D_{i_1}(s) + \lambda D_{i_2}(s)$ where $D_{i_1}(s)$ and $D_{i_2}(s)$ are Kharitonov polynomials for the denominator having either the same even order coefficients or the same odd order coefficients. Similar, if the plant denominator is fixed at some Kharitonov polynomial $D_{i_\bullet}(s)$, we obtain four numerator edges of the form $(1-\lambda)N_{i_1}(s) + \lambda N_{i_2}(s)$ where $N_{i_1}(s)$ and $N_{i_2}(s)$ are Kharitonov polynomials for the numerator having either the same even order coefficients or the same odd order coefficients.

We see below that the λ sweep associated with these edge stability problems can be eliminated when the compensator is first order; i.e. instead of checking 32 edges, one can work with 16 specially designated extreme plants.

Theory for first order compensators

The papers by Hollot and Yang (1990) and Barmish *et al.* (1992) provide an avenue for extreme point analysis in lieu of the 32 edges above. We continue to consider an interval plant family $\mathcal{P} = \{P(\cdot, q, r) : q \in Q; r \in R\}$, assume all members of $\mathcal{P}$ are strictly proper and for simplicity, we take $D(s, r)$ to be monic. Now, with a first order compensator

$$C(s) = K\frac{(s-z)}{(s-p)},$$

as connected as in Fig. 1, we say that $C(s)$ *robustly stabilizes* $\mathcal{P}$ if the resulting family of *closed loop polynomials* described by $q \in Q$, $r \in R$ and

$$p_{\mathrm{CL}}(s,q,r) = K(s-z)N(s,q) + (s-p)D(s,r),$$

is robustly stable. With the setup above, Hollot and Yang (1990) establish the following result: *The first order compensator $C(s)$ robustly stabilizes $\mathcal{P}$ if and only if it stabilizes each of the*

plants $P(s, q^{i_1}, r^{i_2})$ associated with extreme point pairs $(q^{i_1}, r^{i_2}) \in Q \times R$; i.e. for each such pair, the associated closed loop polynomial $p_{CL}(s, q^{i_1}, r^{i_2})$ is stable. Notice that if the plant numerator has degree m and the plant denominator has degree n, this result can involve as many as 2^{m+n+1} extremes. We now describe a stronger result for the identical problem.

The salient feature in the paper of Barmish *et al.* (1992) is a set of 16 *Kharitonov plants* instead of 2^{m+n+1} plants. These 16 plants are obtained by generating the four Kharitonov polynomials $N_1(s)$, $N_2(s)$, $N_3(s)$ and $N_4(s)$ for the numerator and the four Kharitonov polynomials $D_1(s)$, $D_2(s)$, $D_3(s)$ and $D_4(s)$ for the denominator. Subsequently, we obtain 16 distinguished plants defined by

$$P_{i_1, i_2}(s) = \frac{N_{i_1}(s)}{D_{i_2}(s)},$$

for $i_1, i_2 = 1, 2, 3, 4$. The importance of these 16 distinguished plants was first pointed out by Mori and Barnett (1988) and Chapellat and Bhattacharagya (1990) in a rather different context involving unstructured perturbations; see Section 11 for further details.

Theorem 8.1. (Barmish *et al.* (1992)) Consider the strictly proper interval plant family $\mathcal{P}$ with monic denominator and the first order compensator $C(s)$. Then $C(s)$ robustly stabilizes $\mathcal{P}$ if and only if it stabilizes each of the associated 16 Kharitonov plants.

To conclude this subsection, we note that the Theorem 8.1 can be refined when more *a priori* information is assumed about the signs and the magnitudes of K, z and p. For such cases, it can be shown that only eight or 12 plants are required.

Higher order compensators

To see that an extreme point result is not always guaranteed for higher order compensators, we consider the simple interval plant of Hollot and Yang (1990) described by

$$P(s, q) = \frac{}{(s^2 + 25s)},$$

and $q \in [1, 5000]$. Now, with compensator

$$C(s) = \frac{(s+3)(s+4)}{(s+0.1)(s+0.2)(s+75)},$$

connected as in Fig. 1, it is easy to prove that $C(s)$ stabilizes the two extreme plants $P(s, 1)$ and $P(s, 5000)$ but the closed loop system is unstable for $q = 2$. In view of examples of this sort, it is of interest to delineate new compensator classes which lead to extreme point solutions in the interval plant framework. Recalling that we can reduce the robust stability problem to 32 edge problems for compensators of arbitrary order, there is considerable motivation to pursue extreme point results involving convex combination of only two polynomials. We now pursue this topic.

9. CONVEX DIRECTIONS

The results in this section are aimed at enlarging the class of polytopic robust stability problems for which an extreme point result holds. In view of the Edge Theorem, we concentrate on one parameter problems and begin the presentation with a definition due to Rantzer (1991).

Definition (*convex direction*)

A monic polynomial $g(s)$ is said to be a *convex direction* (for the space of stable nth order polynomials) if the following condition is satisfied: Given any stable nth order polynomial $f(s)$ such that $f(s) + g(s)$ is also stable and $\deg(f(s) + \lambda g(s)) = n$ for all $\lambda \in [0, 1]$, it follows that $f(s) + \lambda g(s)$ is stable for all $\lambda \in [0, 1]$.

The convex direction concept is depicted graphically in Fig. 6. From the figure, it is apparent that $g_1(s)$ is a convex direction because $f(s) + \gamma g_1(s)$ remains within the stable set for all $\gamma \geq 0$. On the other hand, $g_2(s)$ is not a convex direction. The important point to note is that convex direction results can readily be reinterpreted in an extreme point context; i.e. if $g(s)$ is a convex direction and $0 \leq \gamma_1 \leq \gamma_2$, then stability of $f(s) + \gamma_1 g(s)$ and $f(s) + \gamma_2 g(s)$ implies stability of $f(s) + \gamma g(s)$ for all $\gamma \in [\gamma_1, \gamma_2]$. We now use the convex direction verbiage for discussion of existing literature.

As seen via a simple example in Bialas and Garloff (1985), not all directions $g(s)$ are

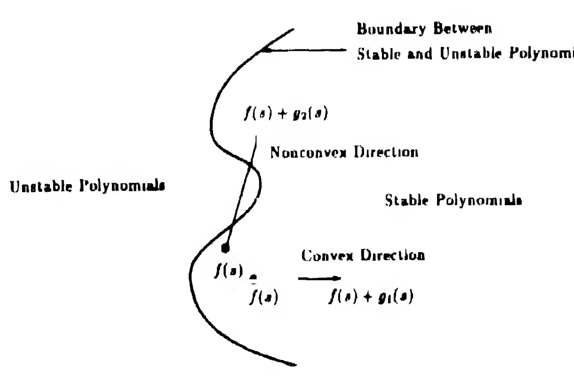


FIG. 6. Convex directions in the space of polynomials.

convex; i.e. if $f(s) = 10s^3 + s^2 + 6s + 0.57$ and $g(s) = s^2 + 2s + 1$, it follows that $f(s)$ and $f(s) + g(s)$ are stable, $f(s) + \lambda g(s)$ has constant degree for all $\lambda \in [0, 1]$ but $f(s) + \frac{1}{2}g(s)$ is unstable. Said another way, Bialas and Garloff (1985) emphasize the well-known fact that a convex combination of stable polynomials need not be stable. This raises the following problem: Provide conditions under which $g(s)$ is a convex direction. An important paper by Rantzer (1992a) provides an elegant solution to this problem. We refer to the condition in the theorem below as *Rantzer's Growth Condition.*

Theorem 9.1. (Rantzer (1992a)) A polynomial $g(s)$ is a convex direction for the space of stable nth order polynomials if and only if

$$\frac{d}{d\omega} \measuredangle g(j\omega) \leq \frac{\sin 2\measuredangle g(j\omega)}{2\omega}$$

for all $\omega > 0$ such that $g(j\omega) \neq 0$.

The theorem above is quite useful for explanation of existing results and delineating new classes of convex combination problems which admit extreme point solution. For example, if we consider the setup in Theorem 9.1 but allow the compensator $C(s) = N_c(s)/D_c(s)$ to be of arbitrary order, the satisfaction of Rantzer's Growth Condition leads to the following result: *Robust stability of* 16 *interconnection pairs* $(P_{i_1,i_2}(s), C(s))$ *corresponding to the Kharitonov plants implies robust stability for all interconnection pairs* $(P(s), C(s))$ *with* $P(s) \in \mathscr{P}$. For example, the compensator

$$C(s) = \frac{s^5 - s^4 - s^3 - s^2 + s - 1}{s^6 - 0.5s^5 + s^4 - 1.5s^3 - 4s^2 - s - 4},$$

is readily verified to have numerator and denominator satisfying the growth condition. Hence, if $C(s)$ is connected to an interval plant as in Fig. 1, robust stability of the closed loop is ascertained by simply checking the stability of 16 closed loop systems.

It is easy to verify that Rantzer's Growth Condition captures earlier extreme point results such as those as given in Rantzer (1990) and Fu (1991). For example, if $\measuredangle g(j\omega)$ is nonincreasing, then the growth condition is trivially satisfied. Of course, since the growth condition is both necessary and sufficient for $g(s)$ to be a convex direction, it also captures (in a less explicit way) other coefficient-based convex direction results in the literature. For the sake of completeness, we review these results because they provide a spectrum of examples of convex directions.

Beginning at the level of interval polynomials, we can interpret the Kharitonov analysis in the convex direction framework by taking $g(s)$ to be a polynomial containing only either all even or all odd powers of s. Such a $g(s)$ turns out to be a convex direction and the associated extreme point result for $f(s) + \lambda g(s)$ can be exploited in a proof of Kharitonov's Theorem. In a similar manner, one can interpret the extreme point results of Bialas and Garloff (1985); i.e. again take $g(s)$ to be a polynomial containing only either all even or all odd powers of s. To recover the result of Petersen (1990), $g(s)$ is taken to be antistable (all roots in the open right half plane) and Rantzer's Growth Condition is verified by inspection. Basic to the paper by Hollot and Yang (1990) is a convex direction of the form $g(s) = s^k(s + \alpha)$ where α is real. A minor extension of this form is exploited by Barmish *et al.* (1992); they work with a convex direction of the form $g(s) = (s + \alpha)h(s)$ with α real and $h(s)$ containing only all even or all odd powers of s.

Motivation for the paper by Barmish and Kang (1992) is derived from the fact that a systematic transcription of the growth condition into a direct condition on the coefficients of $g(s)$ appears to be very complicated. This motivates seeking a simpler convex direction criterion which involves the coefficients of $g(s)$ in a more straightforward way. With this as the goal, the theory of Barmish and Kang begins with a monic polynomial

$$g(s) = s^m + \sum_{i=0}^{m-1} a_i s^i,$$

and the associated *Hurwitz matrix*

$$\mathscr{H}(g) \doteq \begin{bmatrix} a_{m-1} & a_{m-3} & a_{m-5} & \cdots & \cdots \\ 1 & a_{m-2} & a_{m-4} & \cdots & \cdots \\ 0 & a_{m-1} & a_{m-3} & a_{m-5} & \cdots \\ 0 & 1 & a_{m-2} & a_{m-4} & \cdots \\ 0 & 0 & a_{m-1} & a_{m-3} & \cdots \\ 0 & 0 & 1 & a_{m-2} & \cdots \\ 0 & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdot & a_4 & a_2 & a_0 \end{bmatrix}$$

Taking $\Delta_1(g), \Delta_2(g), \ldots, \Delta_m(g)$ to be the leading principal minors of $\mathscr{H}(g)$, we say that $g(s)$ satisfies the *Alternating Hurwitz Minor Condition* (AHMC) if at least one of the following two conditions hold.

(i) The associated odd order Hurwitz minors of $g(s)$ alternate according to the sign pattern

$$\Delta_1(g) < 0, \quad \Delta_3(g) > 0,$$
$$\Delta_5(g) < 0, \quad \Delta_7(g) > 0, \ldots.$$

(ii) The associated even order Hurwitz minors of $g(s)$ alternate according to the sign pattern

$$\Delta_2(g) < 0, \quad \Delta_4(g) > 0,$$
$$\Delta_6(g) < 0, \quad \Delta_8(g) > 0, \ldots .$$

For the degenerate case when $m = 1$, we adopt the convention that (ii) holds vacuously. Finally, if for some $k \geq 0$, either condition (i) or condition (ii) holds except for the fact that $\Delta_i(g) = 0$ for $i \geq k$, then $g(s)$ is said to satisfy the *extended AHMC*. A simple convex direction condition can now be given.

Theorem 9.2. (Barmish and Kang (1992)) Suppose $g(s)$ is a monic polynomial of degree m which satisfies the extended AHMC. Then $g(s)$ is a convex direction for the space of polynomials of order $n > m$.

The application of the AHMC to SISO feedback systems leads to a number of extreme point results. For example, it is easy to verify that numerator and denominator of the compensator

$$C(s) = \frac{10(s+4)(s+1)(s-2)}{(s^2-3s+4)(s+2)(s-3)},$$

satisfy AHMC conditions (ii) and (i), respectively. Therefore, if the plant is strictly proper with monic denominator, it is easy to verify that $C(s)$ robustly stabilizes an interval plant family $\mathscr{P}$ if and only if it stabilizes the associated 16 Kharitonov plants.

10 RESULTS AT THE MATRIX LEVEL

The motivation for this section is derived from state space systems which are described in terms of the vector of uncertain parameters q; we consider

$$\dot{x}(t) = A(q)x(t),$$

and call $A(q)$ an *uncertain matrix*. In view of the discussion in the preceding sections, there are a number of "quick and dirty" results which can be obtained by noting that the polynomial

$$p(s, q) = \det(sI - A(q)),$$

governs the robust $\mathscr{D}$-stability for the family of matrices $\mathscr{A} = \{A(q) : q \in Q\}$. For example, suppose $A(q)$ is a companion canonical form; e.g. say

$$A(q) = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ q_0 & q_1 & q_2 & \cdots & q_{n-1} \end{bmatrix}$$

Then, we obtain a simple matrix analogue of Kharitonov's Theorem: *$\mathscr{A}$ is robustly stable if and only if four distinguished extreme matrices are stable.* Of course, the four distinguished matrices are obtained from the Kharitonov polynomials associated with the interval polynomial family $\mathscr{P} = \{p(\cdot, q) : q \in Q\}$.

There are also some trivial matrix analogues of robust stability results for polytopes of polynomials. For example, if q enters affine linearly into only a single row or column of $A(q)$, then the resulting characteristic polynomial $p(s, q)$ turns out to have an affine linear uncertainty structure and many results already surveyed can be restated in a matrix setting.

In fact, there is a class of rank one matrix uncertainty structures which permit linkage with the theory for polytopes of polynomials. This class is nicely characterized in the dissertation by El Ghaoui (1990). Rewriting $A(q)$ in the form

$$A(q) = A_0 + \sum_{i=0}^{\ell} A_i q_i,$$

with A_i fixed for $i = 0, 1, 2, \ldots, \ell$, the following result holds: *The characteristic polynomial $p(s, q) = \det(sI - A(q))$ has affine linear uncertainty structure if and only if*

$$\min\{\operatorname{rank}[A_1, A_2, \ldots, A_\ell],$$
$$\operatorname{rank}[A_1', A_2', \ldots, A_\ell']\} = 1,$$

where A_i' denotes the transpose of A_i.

Interval matrices

In this subsection, we work with matrices having uncertain parameters each entering into one entry at most. For convenience, we use a double subscripting notation and denote the (i, j)th entry of $A(q)$ by q_{ij}. This enables us to use a compact notation to describe an *interval matrix family*, i.e. if $[q_{ij}^-, q_{ij}^+]$ is a given bound for q_{ij}, we can use a notation similar to the one employed for interval polynomials. For example, we can describe a 2×2 interval matrix family by writing

$$A(q) = \begin{bmatrix} [q_{11}^-, q_{11}^+] & [q_{12}^-, q_{12}^+] \\ [q_{21}^-, q_{21}^+] & [q_{22}^-, q_{22}^+] \end{bmatrix}.$$

For an interval matrix family $\mathscr{A}$, we let A_i denote the ith extreme matrix; e.g. if q^i is the ith extreme point of Q, we can take $A_i \doteq A(q^i)$. The first point to mention is that for interval matrices, extreme point results for robust stability are the exception rather than the rule. For example, the 3×3 interval matrix of

Barmish and Hollot (1984)

$$A(q) = \begin{bmatrix} [-1.5, -0.5] & -12.06 & -0.06 \\ -0.25 & 0 & 1 \\ 0.25 & -4 & -1 \end{bmatrix}.$$

has stable extremes A_1 and A_2 but is unstable when $q_{11} = -1$; see also Karl *et al.* (1984) for a similar example.

It is interesting to note that a 3×3 matrix was used above. At the level of 2×2 matrices, if we use the fact that stability is equivalent to positivity of the coefficients of the associated second order characteristic polynomial, we can easily obtain an extreme point result; e.g. in Mansour (1989) it is seen that stability of the set of 2×2 extreme matrices $\{A_i\}$ is equivalent to robust stability of $\mathscr{A}$. As far as 3×3 interval matrices are concerned, the extreme point result of Zong (1990) can be applied provided the diagonal elements of $\mathscr{A}$ are fixed; i.e. if $q_{ii}^- = q_{ii}^+$ for $i = 1, 2, 3$, then robust stability of the extreme matrices $\{A_i\}$ is equivalent to robust stability of $\mathscr{A}$. At the level of 4×4 matrices and above, extreme point results are available only under rather strict conditions.

Polytopes of matrices and Lyapunov functions

If the entries of $A(q)$ depend affine linearly on q, then the matrix family $\mathscr{A} = \{A(q) : q \in Q\}$ is called a *polytope of matrices* or a *polytopic matrix family*. The fact that $\mathscr{A}$ is polytopic is explained by noting that if q^i is the ith extreme point of Q and again, we take $A_i \doteq A(q^i)$, then

$$\mathscr{A} = \text{conv}\,\{A_i\}.$$

Since the interval matrix problem is a special case of the polytope of matrices problem, we cannot expect extreme point results at the level of 4×4 and above without enforcing additional conditions. Since such conditions are quite restrictive, we only provide a sampling of the available results along these lines.

The following result is given by Shi and Gao (1986): If *$\mathscr{A}$ is a polytope of symmetric matrices, then stability of the set of extreme matrices $\{A_i\}$ is equivalent to robust stability of $\mathscr{A}$*. This is easily established by observing that if the extremes are stable, then with $P = P' = I$, we have $A_i'P + PA_i < 0$. That is, the identity serves as a *common Lyapunov matrix* for the extremes. Using the fact that positively weighted sums of negative–definite matrices are still negative–definite and noting that every $A \in \mathscr{A}$ is a convex combination of the A_i, it follows that $A'P + PA < 0$ for all $A \in \mathscr{A}$. Hence $P = I$ serves as a common Lyapunov matrix for all of $\mathscr{A}$; e.g. see Fu (1987).

The usefulness of common Lyapunov functions in an extreme point context is not particular to the symmetric matrix analysis. Notice that we can easily modify the argument above so that it applies to a general polytope of matrices; e.g. in the analysis of Horisberger and Belanger (1976), we see that if a polytope of matrices $\mathscr{A}$ (not necessarily symmetric) admits common Lyapunov matrix $P = P' > 0$ for its extremes, this same matrix serves for all of $\mathscr{A}$. Furthermore, it can also be shown that the problem of finding a common Lyapunov matrix for the extremes is a convex program; see Boyd and Barratt (1990) for further details.

More generally, if $\mathscr{A} = \{a(q) : q \in Q\}$ is a family of matrices with entries depending multilinearly on q, similar extreme point results are obtained. Along these lines, the result of Garofalo *et al.* (1992) is the most general: *If each entry of $A(q)$ is a ratio of multilinear functions of q, then once again, a common Lyapunov matrix for the extremes of Q serves as a common Lyapunov matrix for all of $\mathscr{A}$.*

Function maximization on matrix polytopes

In some applications, we begin with a polytope of matrices $\mathscr{A} = \{A(q) : q \in Q\}$ and a convex function $f : \mathscr{A} \to \mathbf{R}$ which represents some aspect of performance. Using the basic fact that a convex function on a polytope is maximized at an extreme point, we conclude that

$$\max_{A \in \mathscr{A}} f(A) = \max_{i} f(q^i).$$

Using this simple idea, a number of interesting "small" results can be obtained. For example, if $\mathscr{A}$ is a symmetric polytope of matrices, we can reproduce the extreme point results given in Section 10. These ideas also lead to a simple proof of a result given in Mori and Kokame (1987); i.e. let $\|\cdot\|$ be any matrix norm, take $\gamma > 0$ as given and consider an interval matrix family $\mathscr{A} = \{A(q) : q \in Q\}$. Then using $f(A) = \|A\|$ above, it follows that $\|A\| < \gamma$ for all $A \in \mathscr{A}$ if and only if $\|A(q^i)\| < \gamma$ for all extreme points q^i of Q.

Robust nonsingularity

Another interesting set of results at the matrix level exploits a certain relationship between robust stability and *robust nonsingularity*. To clearly explain this linkage, we take $\mathscr{A} = \{A(q) : q \in Q\}$ be a family of $n \times n$ matrices with $A(q)$ depending continuously on q. Under rather mild conditions, there is a family of linear transformations (for example, see Bialas (1985) and Fu and Barmish (1988)) mapping $\mathscr{A}$ into a new family $\bar{\mathscr{A}}$ having the following property. *The*

family $\mathscr{A}$ is robustly stable if and only if the family $\tilde{\mathscr{A}}$ is robustly nonsingular; i.e. all matrices $\tilde{A} \in \tilde{\mathscr{A}}$ are nonsingular.

One well-known example of such a transformation involves the Kronecker operations. Indeed, if A and B are square matrices of dimension n_1 and n_2, respectively and a_{ij} is the (i, j)th entry of A, the *Kronecker product* $A \otimes B$ is a square matrix of dimension $n_1 n_2$ with (i, j)th block $a_{ij}B$. Subsequently the *Kronecker sum* is defined by

$$A \otimes B \doteq A \oplus I_{n_2} + I_{n_1} \otimes B,$$

where I_{n_1} and I_{n_2} denote identity matrices of dimensions n_1 and n_2, respectively. With this notation, we define a linear transformation T on a polytope of matrices $\mathscr{A}$ as follows: If $a \in \mathscr{A}$, then

$$TA \doteq A \oplus A.$$

The important point to note is that any linear transformations on $\mathscr{A}$ preserves the affine linear dependence of matrix entries on the uncertain parameters.

For the reader interested in further details on Kronecker operations, the review paper by Brewer (1978) is recommended. We now present a result which applies to nonsingularity of interval matrices. Using the basic fact that a multilinear function on a box achieves both its maximum and minimum at an extreme point, the following lemma is easy to prove by noting that $\det A$ depends multilinearly on the entries of A.

Lemma 10.1. Suppose $\mathscr{A} = \{A(q) : q \in Q\}$ is an interval matrix family. Then $\mathscr{A}$ is robustly nonsingular if and only if each of the determinants $\det A(q^i)$ has the same sign.

A weakness of this result is the "combinatoric explosion" associated with the solution. To illustrate what is meant by this, notice that if $\mathscr{A}$ is an $n \times n$ interval matrix family, there can be as many as 2^{n^2} extreme points. The results of Rohn (1989) enables us to reduce this number to 4^n.

Rohn's framework

We begin with a nonsingular $n \times n$ matrix A_0 and given bounds $r_{ij} \geq 0$ for the entries q_{ij} of an interval matrix $\Delta A(q)$. We now define the family of matrices

$$\mathscr{A}_r \doteq \{A_0 + \epsilon \Delta A(q) : 0 \leq \epsilon \leq r, q \in Q\},$$

with variable *magnification factor* $r \geq 0$. The objective is to obtain the robustness margin

$$r_{\max} \doteq \sup\{r : \mathscr{A}_r \text{ is robustly nonsingular}\}.$$

Taking R to be the $n \times n$ matrix having (i, j)th entry r_{ij}, we provide some definitions needed in order to describe Rohn's result; see also Demmel (1988). Indeed, if M is a square matrix, let

$$\rho_0(M) \doteq \max\{|\lambda| : \lambda \text{ is a real eigenvalue of } M\},$$

with $\rho_0(M) \doteq 0$ if no eigenvalues of M are real. A square matrix S is said to be a ***signature matrix*** if it is diagonal with all diagonal entries equal to either $+1$ or -1. Now, we let $\mathscr{S}$ be the set $n \times n$ signature matrices and observe that $\mathscr{S}$ has 2^n members. We are now prepared to present Rohn's Theorem.

Theorem 10.2. (Rohn (1989)) Given $r \geq 0$, the interval matrix family $\mathscr{A}_r$ is robustly nonsingular if and only if $\det A_0$ and the 4^n determinants $\det(A_0 + rS_1RS_2)$, obtained with $S_1, S_2 \in \mathscr{S}$, have the same nonzero sign. Moreover,

$$r_{\max} = \frac{1}{\max\limits_{S_i \in \mathscr{S}} \rho_0(S_1 A_0^{-1} S_2 R)}.$$

Connection with μ theory

In this brief subsection, we mention an interesting connection between Rohn's result and the computation of the real structured singular value (real μ) for the case when the ℓ^∞ norm is used for the uncertain parameter vector q. To this end, we begin with a real square rank one matrix M and the goal is to compute

$$\frac{1}{\mu_\infty(M)} \doteq \inf\{\|q\|_\infty : \det(I + M\Delta(q)) = 0\},$$

where $\|q\|_\infty$ denotes the ℓ^∞ norm of q and

$$\Delta(q) \doteq \operatorname{diag}\{q_1, q_2, \ldots, q_\ell\}.$$

In the numerical computation of $\mu_\infty(M)$, the following lower bound has traditionally been used:

$$\mu_\infty(M) \geq \max_{S \in \mathscr{S}} \rho_0(SM).$$

In view of the extreme point theory of Rohn (1989), it can be shown that the lower bound is sharp. For further discussion of extreme point results in this framework, see also Holohan and Safonov (1989), El Ghaoui (1990) and Chen *et al.* (1992).

Stability and Schur stability of non-negative interval matrices

In this subsection, we survey some extreme point results on stability and Schur stability of interval matrices with non-negative entries. The motivation for this problem is derived from a number of different fields. For example, this

problem arises if we consider the transition matrix of a Markov chain with non-negative switching probabilities only known within given bounds. We begin the exposition by taking A to be an $n \times n$ square matrix with non-negative off-diagonal entries. Letting A_k denote the upper $k \times k$ block of A for $k = 1, 2, \ldots, n$, in accordance with classical results from matrix algebra (for example, see Gantmacher (1959)), it follows that A is stable if and only if

$$(-1)^k \det (I_k - A_k) > 0,$$

for $k = 1, 2, \ldots, n$ where I_k denotes an identity matrix of dimension k. A similar result holds for Schur stability with the added restriction that the diagonal entries of A are non-negative and the determinant condition above is replaced by $\det (I_k - A_k) > 0$.

If we consider an $n \times n$ interval matrix family $\mathcal{A} = \{A(q) : q \in Q\}$ with non-negative off-diagonal entries, we can exploit the result above in combination with the arguments immediately preceding Lemma 10.1 to arrive at the following extreme point result: *$\mathcal{A}$ is robustly stable if and only if*

$$(-1)^k \det (I_k - A_k(q^i)) > 0,$$

for all extreme points q^i of Q and all $k \in \{1, 2, \ldots, n\}$. Furthermore, using ideas quite similar to those in Theorem 10.2, the number of critical extreme points can be dramatically reduced. For the case of robust Schur stability, minor modifications of the arguments above lead to a similar result given by Shafai *et al.* (1991). For further extensions involving irreducible interval matrices, see Mayer (1984).

11. OTHER ASPECTS OF PERFORMANCE

Over the last few years, we see the beginning of a collection of extreme point results involving other aspects of robust performance besides robust $\mathcal{D}$-stability. The first result which we mention involves H^∞ analysis with structured real uncertainty. For completeness, recall that if $P(s)$ is proper, stable and rational, then the H^∞ norm is given by

$$\|P\|_\infty \doteq \sup_\omega |P(j\omega)|.$$

Theorem 11.1. (See Mori and Barnett (1988) and Chapellat *et al.* (1990)) Consider a robustly stable proper interval plant family $\mathcal{P} = \{P(s, q, r) : q \in Q, r \in R\}$ with monic denominator and 16 associated Kharitonov plants $P_{ij}(s)$ for $i, j = 1$–4. Then it follows that

$$\max_{q \in Q, r \in R} \|P(\cdot, q, r)\|_\infty = \max_{i,j} \|P_{ij}\|_\infty.$$

The ideas central to the proof of Theorem 11.1 enter into the proof of a number of closely related results. For example, in Chapellat *et al.* (1990), a real parameter version of the Small Gain Theorem is established and in Mori and Barnett (1988), a robust Popov-like criterion is provided in the interval plant context. These results are extended in Dahleh *et al.* (1991), Vicino and Tesi (1991) and Rantzer (1992a). It is also interesting to point out that the results of the sort above are not readily extendable to a more general framework involving weighting functions or restricted frequency intervals; e.g. see the paper by Hara *et al.* (1991) for further discussion.

Positive-realness

The issue of positive-realness (addressed in the classical Popov theory) is studied in an extreme point context by Dasgupta and Bhagwat (1987). If $P(s, q) = N(s, q)/D(s, q)$ is an uncertain plant and q is fixed, we say that this transfer function is strictly positive real (SPR) if both $N(s, q)$ and $D(s, q)$ are stable and $\operatorname{Re} P(j\omega, q) > 0$ for all $\omega \in \mathbf{R}$. We say that a family of plants $\mathcal{P} = \{P(\cdot, q) : q \in Q\}$ is *robustly* SPR if $P(s, q)$ is SPR for all $q \in Q$.

Theorem 11.2. (Dasgupta and Bhagwat (1987)) Consider the interval plant family $\mathcal{P}$ with fixed numerator $N(s, q) \equiv N(s)$. Letting $D_1(s)$, $D_2(s)$, $D_3(s)$ and $D_4(s)$ denote the four Kharitonov denominator polynomials, it follows that $\mathcal{P}$ is robustly SPR if and only if $N(s)/D_i(s)$ is SPR for $i = 1$–4.

In their paper, Dasgupta and Bhagwat also point out the applicability of these results in an adaptive output error identification context. There are also a number of papers dealing with extensions and variations on the SPR theme. In papers by Dasgupta (1987), Bose and Delanski (1989), Chapellat *et al.* (1991) and Shi (1991), the SPR problem is studied under the weaker hypothesis that the numerator $N(s, q)$ is uncertain as well; in some cases, plants with complex coefficients are considered. It is also worth noting that extreme point results for SPR problems with multilinear uncertain structures are also available. However, instead of a four plant result as in Theorem 11.2, the plants associated with all of the extreme q^i come into play; see the recent paper by Dasgupta *et al.* (1991). Finally, we mention extreme point results for the case when the plant is not necessarily SPR but can be rendered SPR by suitable addition of a positive constant; see Chapellat *et al.* (1991).

Steady state and overshoot

In the work of Bartlett (1990), an uncertain plant $P(s, q)$ with multilinear uncertainty structure is considered and steady state error for a unit step input is the prime consideration. A robust stability assumption is imposed and it is shown that *both the maximum and the minimum of the steady state error occur on one of the extreme plants* $P(s, q^i)$. Bartlett then goes on to provide a counterexample to the tempting conjecture that the maximal peak overshoot is also attained at an extreme. To this end, he considers the family of plants with affine linear uncertainty structure described by

$$P(s, q) = \frac{1}{(3.4q + 0.1)s^2 + (1.7q + 0.8)s + 1},$$

and $q \in [0, 1]$. It turns out that the peak overshoot corresponding to a step input is not maximized at the extremes $q = 0$ or $q = 1$; e.g. $q = 0.5$ leads to a higher overshoot value. A similar example is also given for discrete time systems.

The Nyquist envelope

To complete this section, we briefly discuss an emerging line of work related to the frequency response of an interval plant. Indeed, suppose that $\mathcal{P} = \{P(s, q, r): q \in Q, r \in R\}$ is a strictly proper interval plant family with monic denominator and, to keep the exposition simple, we also assume that for all $r \in R$, the plant denominator $D(s, r)$ has no roots on the imaginary axis. Then, the *Nyquist set* associated with this family of plants is defined by

$$\mathcal{N} \doteq \{P(j\omega, q, r): \omega \in \mathbf{R}; q \in Q; r \in R\}.$$

In Hollot and Tempo (1991), the boundary $\partial\mathcal{N}$ of the Nyquist set, called the *Nyquist envelope*, is the focal point and the following question is addressed: What points $z \in \partial\mathcal{N}$ on the Nyquist envelope are Kharitonov points? By this, we mean points $z \in \partial\mathcal{N}$ such that $z = P_{i_1,i_2}(j\omega)$ for one of the 16 Kharitonov plants $P_{i_1,i_2}(s)$ and some $\omega \in \mathbf{R}$. If $\mathcal{P}$ is robustly stabilized by unity feedback, then the distinguished points on the envelope include minimal gain margin points and minimal phase margin points. If $P(s, q)$ is (open loop) stable for all $q \in Q$, the points associated with the maximum H^∞ norm also lie on the Nyquist envelope. Such a point is identified with a $q^* \in Q$ and an $\omega^* \in \mathbf{R}$ such that

$$|P(j\omega^*, q^*)| = \max_{q,\omega} |P(j\omega, q)|.$$

Recalling Theorem 11.1 and the discussion in Section 5, we already know that the gain margin point and the H^∞ point are Kharitonov points.

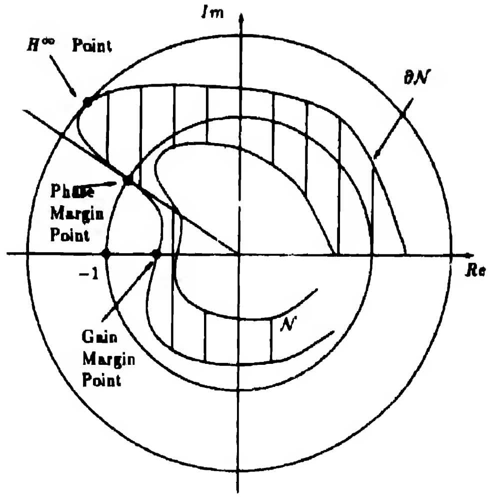


FIG. 7. Gain margin, phase margin and H^∞ points are Kharitonov points.

The paper by Hollot and Tempo (1991) provides a characterization of a class of Kharitonov points via a certain "path condition". Their theory identifies some new Kharitonov points and recovers the ones which we already know. The fact that critical points associated with minimum phase margin, sensitivity and complementary sensitivity are identified as Kharitonov points seems to be new. In this regard, see also the paper by Kimura and Hara (1991) where a more restrictive class of interval plants with fixed numerator is considered. To further clarify the geometry associated with the discussion above, we refer to Fig. 7 where three distinguished points on the Nyquist envelope are depicted.

12. RESULTS FOR OTHER UNCERTAINTY REPRESENTATIONS

Throughout this paper, it has been assumed that the uncertainty bounding set Q is a box. In this section, we consider two alternative descriptions of Q. In each case, new extreme point results are obtained.

The rotated rectangle for robust Schur stability

The motivation for the exposition in this subsection is derived from the fact that stringent assumptions are imposed in Section 6 in order to obtain extreme point results for robust Schur stability of an interval polynomial family. In the work of Kraus *et al.* (1988b), a new uncertainty model is considered. These authors begin with an uncertain polynomial $p(s, q) = \sum_{i=0}^{n} q_i s^i$ having independent uncertainty structure but dispose with the box bound Q for q. Instead, they consider a rotated rectangular bound of the sort shown in Fig. 8; other quadrants for this

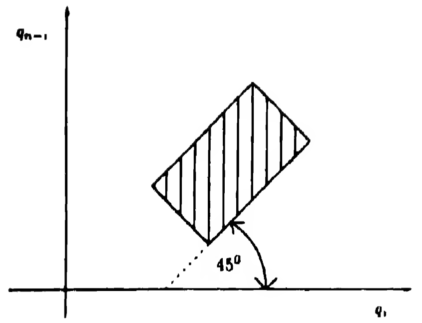


FIG. 8. Bounding set for the pair (q_i, q_{n-i}).

rectangle are also permitted. For the case when n is even and $i = n/2$, an interval $[q^-_{n/2}, q^+_{n/2}]$ is taken as a bound for $q_{n/2}$. When the uncertain parameters are bounded via a set Q in the manner above, we say that Q is a *product of* 45° *rotated rectangles*. As usual, we let $\{q^i\}$ denote the set of extreme points of Q. With this new setup, rather strong extreme point result is given below.

Theorem 12.1. (Kraus *et al.* (1988b)) Consider the uncertain family of polynomials $\mathcal{P} = \{p(\cdot, q): q \in Q\}$ with invariant degree and Q being a product of 45° rotated rectangles. Then, it follows that $\mathcal{P}$ is robustly Schur stable if and only if $p(s, q^i)$ is Schur stable for all extreme points q^i of Q.

In a follow-up paper by Mansour *et al.* (1988) even stronger results are established. Under the hypotheses of the theorem above, they show that a "small" subset of the extreme polynomials need only to be considered. Furthermore, working with Chebyshev and Jacobi polynomials, they develop a recipe for selection of these distinguished extremes. More recently, a novel technical method (involving barycentric coordinates) of Perez *et al.* (1992) has resulted in extreme point results for classes rotated rectangles other than 45°. For angular rotations in the sector $[\pi/4, 3\pi/4]$, results along the lines of Theorem 12.1 are given.

Diamond families of polynomials

In a paper by Tempo (1990), a robust stability problem is formulated which can be viewed as a dual to the problem of Kharitonov. Whereas the interval polynomial framework is associated with the ℓ^∞ norm for q, Tempo considers the ℓ^1 norm and works with an uncertain polynomial $p(s, q) \doteq \sum_{i=0}^{n} q_i s^i$ with independent uncertainty structure and uncertainty bounding set Q to be the *diamond* described by

$$\sum_{i=0}^{n} |q_i - q_i^*| \le r.$$

We call $p(s, q^*) = \sum_{i=0}^{n} q_i^* s^i$ the *center polynomial* and refer to $\mathcal{P} = \{p(\cdot, q): q \in Q\}$ as a *diamond family of polynomials*. Associated with this family $\mathcal{P}$ is a set of eight *critical polynomials* defined by

$$\begin{aligned}
c_1(s) &\doteq p(s, q^*) + r;\\
c_2(s) &\doteq p(s, q^*) - r;\\
c_3(s) &\doteq p(s, q^*) + rs;\\
c_4(s) &\doteq p(s, q^*) - rs;\\
c_5(s) &\doteq p(s, q^*) + rs^{n-1};\\
c_6(s) &\doteq p(s, q^*) - rs^{n-1};\\
c_7(s) &\doteq p(s, q^*) + rs^{n};\\
c_8(s) &\doteq p(s, q^*) - rs^{n}.
\end{aligned}$$

In the theorem below, we see that the critical polynomials $c_i(s)$ play the same role in the ℓ^1 analysis as the Kharitonov polynomials $K_i(s)$ play in the ℓ^∞ analysis.

Theorem 12.2. (Barmish *et al.* (1990)) A diamond family of polynomials $\mathcal{P}$ having invariant degree is robustly stable if and only if the associated eight critical polynomials $c_1(s), c_2(s), \ldots, c_8(s)$ are stable.

Unlike Kharitonov's Theorem, the extreme point results for the ℓ^1 case above do not easily extend to weighted norms. To demonstrate this point, we consider the weighted diamond family of polynomials in the dissertation of Kang (1992): The *center polynomial* is given by $p(s, q^*) = s^3 + 2s^2 + 2.201s + 4$ and admissible coefficients are described by $10|q_3 - q_3^*| + 100|q_2 - q_2^*| + 100|q_1 - q_1^*| + 2.5|q_0 - q_0^*| \le 1$. It turns out that all extreme polynomials $p(s, q^i)$ are stable but the polynomial $\bar{p}(s) = 1.05s^3 + 2s^2 + 2.201s + 4.2$ is an unstable member of this diamond family.

We now describe another issue associated with the ℓ^1 case. When dealing with a complex coefficient polynomials, there are different ways to describe the diamond bound Q. For example, do we formulate the problem with separate diamond bounds for real and imaginary parts or with a "joint" bound? For the case of separate diamond bounds for real and imaginary parts, Barmish *et al.* (1990) consider a second order complex coefficient diamond family described by $p(s, q) = q_2 s^2 + q_1 s + q_0$ with complex uncertain parameters $q_i = u_i + jv_i$ for $i = 0, 1, 2$ and center

polynomial

$$p(s, u^*, v^*) = (1.2272 + j6.3118)s^2 + (0.0111 + j15.1285)s + (-4.3176 + j1.8398).$$

Using diamond bounds $|u_0 - u_0^*| + |u_1 - u_1^*| + |u_2 - u_2^*| \leq 1$ and $|v_0 - v_0^*| + |v_1 - v_1^*| + |v_2 - v_2^*| \leq 1$ for u and v, it is straightforward to verify that all extremes are stable but the polynomial $\bar{p}(s) = p(s, u^*, v^*) + 0.5s - (0.5 + j)$ is an unstable member of this family. On the other hand, if we describe a diamond family by the *joint bound*

$$\tfrac{1}{2}|u_0 - u_0^*| + \tfrac{1}{2}|v_0 - v_0^*| + \sum_{i=1}^{n-1} (|u_i - u_i^*| + |v_i - v_i^*|) \leq r,$$

the extreme point result of Tempo (1989) applies for the case of robust Schur stability.

Markov parameters and other transformations

Rather than using a box to bound the coefficients, the paper by Hollot (1989) considers an uncertainty representation in the space of Markov parameters. To describe the unique correspondence between a fixed polynomial $p(s) = \sum_{i=0}^{n} a_i s^i$ and its Markov parameters, we first break $p(s)$ into even and odd parts as

$$p(s) = f(s^2) + sg(s^2),$$

and for simplicity, we take $f(\cdot)$ and $g(\cdot)$ to be coprime. Now, it is easy to obtain the *Markov parameters* b_i via the expansion

$$\frac{g(x)}{f(x)} = b_{-1} + \frac{b_0}{x} - \frac{b_1}{x^2} + \frac{b_2}{x^3} - \cdots.$$

If $n = 2m$ is even, we take $(b_0, b_1, \ldots, b_{2m-1})$ as the Markov parameters and if $n = 2m - 1$ is odd, we take $(b_{-1}, b_0, \ldots, b_{2m-1})$. Note that Markov parameters can be calculated in a variety of ways; e.g. see Gantmacher (1959). Hollot's paper demonstrates the usefulness of Markov's Theorem in a robust stability context. Namely, if polynomial uncertainty is represented by a box B in the space of Markov parameters, robust stability is guaranteed if and only if two distinguished polynomials are stable. For example, if $n = 2m$ is even and B described by $b_i^- \leq b_i \leq b_i^+$ for $i = 0, 1, 2 \ldots, 2m - 1$, the first distinguished polynomial has Markov parameters $(b_0^-, b_1^+, b_2^-, \ldots, b_{2m-1}^+)$ and the second distinguished polynomial has Markov parameters $(b_0^+, b_1^-, b_2^+, \ldots, b_{2m-1}^-)$.

The results presented in this section address robust stability analysis with the following idea in mind. *An alternative uncertainty representation in a transformed domain often leads to extreme point results which are not possible in the original domain.* Of course, a fundamental limitation of these results is that the relationship between the transformed parameters and the original parameters may be quite complicated. There are a number of other papers involving the transformation point of view. For example, in Petersen (1989), the Delta transform (for the discrete time systems) is implicit in his analysis of $\mathcal{D}$ regions which are shifted circles. For similar extreme point results involving Delta transformation, see the papers by Soh (1991) and Hirsh (1992). Finally, we mention the paper by Vaidyanathan (1990) which considers yet a different transformation for discrete time and the paper by Wei and Yedavalli (1987) which deals with the continuous time case.

Moving away from univariate polynomials

We briefly mention some research aimed at the attainment of extreme point results for delay systems and distributed systems. The paper by Hu *et al.* (1991) raises the question of extreme point results for robust stability of sampled data control systems. They show that the road is rocky via the simple example involving an uncertain plant described by $P(s, q) = 1/(s + q)$ and $q \in [-0.5, 1]$. They close the loop with unity feedback and A/D and D/A converters on each side of $P(s, q)$ and arrive at the uncertain characteristic polynomial

$$h(z, q) = \begin{cases} z + 1 - e^{-qT} - qe^{-qT} & \text{if } q \neq 0, \\ z + T - 1 & \text{if } q = 0, \end{cases}$$

where T is the sampling period. It is then readily verified that for $T = 2.1$, the two extreme polynomials $h(z, -0.5)$ and $h(z, 1)$ are Schur stable but the intermediate polynomial $h(z, 0)$ is not Schur stable.

Counterexamples of the sort above can also be given for delay systems having open loop description $P_T(s, q, r) = P(s, q, r)e^{-sT}$ with $T > 0$ fixed and $P(s, q, r)$ being an uncertain plant with independent uncertainty structure and interval bounds for the components of q and r; e.g. see Ackermann *et al.* (1990). In the same paper extreme point results are given under further restrictions on the plant. For example, if the plant denominator is fixed, the plant numerator is an interval polynomial and unity feedback is applied, closed loop stability can be ascertained by working with the extreme points q^i of Q. The interested reader should also consult the paper by Kim and Bose (1990) where, in a similar vein, the issue of extreme

point results is addressed for classes of delay systems satisfying various technical assumptions.

To conclude this section, we mention a body of work aimed at extension of Kharitonov's Theorem to *scattering Hurwitz polynomials*. In the papers by Bose (1988), Kim and Bose (1988) and Basu (1989), the uncertain polynomial $p(s, q)$ is replaced by a multivariate polynomial $p(s_1, s_2, \ldots, s_n, q)$ and interval bounds on the coefficients are imposed. These papers go on to develop extreme point results which appear to be more relevant to signal processing than control.

13. CONCLUSIONS

The theory of robust stability for polytopes of polynomials seems to have reached a mature state. For polytopes of matrices, however, much research remains to be done. An interesting result illustrating the difficulty of this problem is given in the paper by Cobb and DeMarco (1989): *If $\mathscr{A}$ is a polytope of $n \times n$ matrices with $n \geq 3$, then stability of all faces of dimension $2n - 4$ is sufficient to guarantee robust stability of $\mathscr{A}$. Moreover, if the dimension of $\mathscr{A}$ (viewed as a subset of $\mathfrak{R}^{n^2}$) is $2n - 4$ or greater, there are examples of matrix polytopes which are unstable but have the property that all faces of dimension $2n - 5$ are stable.* Results of this type motivate further research involving the computational complexity of robust stability problems. Some initial results in this direction are given in the paper by Rohn and Poljak (1992) where a class of robust nonsingularity problems are shown to be NP-hard. Another early result is given in Coxson and DeMarco (1992) where it is shown that the problem of deciding if the real structured singular value of a matrix is bounded above by a given constant is NP-hard.

Given the fact that extreme point theory applies to rather specialized situations, the following question is important: What role will extreme point results have (if any) in an "ultimate" robustness analysis theory which is yet to emerge? On one hand, researchers working towards extreme point solutions fully recognize that their "elegant" solutions apply to a very limited problem class. On the other hand, if one attacks a robustness analysis problem with "real world specifications", one might quickly become involved in a whole host of issues ranging from problems involving distinction between local and global minima to high computational complexity.

The perspective of the authors of this survey is that a good direction for future research is described by the words *tailored algorithm*. We elaborate on this point by first describing an "untailored" algorithm: Suppose one takes a robustness analysis problem, "blindly" massages it into a mathematical programming problem and then applies some software package. Then, we call such an algorithm *untailored* in the sense that the underlying control structure is being totally ignored. As an alternative, however, there is a possibility of tailoring the iterative steps of the algorithm by exploiting analytical results which apply to some approximate version of the problem at hand. This idea is well illustrated by the line of work beginning in Saeki (1986) and continued by authors such as de Gaston and Safonov (1988) and Sideris and Sanchez Pena (1989). In these papers the plant numerator and denominator have *multilinear uncertainty structures* but using the Mapping Theorem (see Zadeh and Desoer (1963)), it becomes possible to use results on affine linear uncertainty structures at each step in the iterative process. In other words, these authors provide tailored algorithms which exploit analytical results which apply to an idealization of the problem at hand. Roughly speaking, the error between the solution to the idealized problem and the solution to the original problem is corrected by the iterative process.

To further illustrate the notion of tailored algorithms, we consider analytical results associated with μ theory. The underlying control system structure sometimes makes it possible to generate convex bounds for the solution which would not be ordinarily available for an arbitrary mathematical program. The systematic exploitation of such bounds leads to tailored algorithms. For example, when minimizing some function $f(x)$, the clever exploitation of a convex lower bound $f_{LB}(x)$ might greatly facilitate numerical computation.

Acknowledgements—The authors wish to express their gratitude to Jerry Hamann, Thomas Higgins, Christopher Hollot, Jacob Kogan, Anders Rantzer and Roberto Tempo. Their valuable comments provided during the evolution of this manuscript were most helpful. In addition, the constructive comments provided by the anonymous reviewers are also gratefully acknowledged.

REFERENCES

Ackermann, J. E. (1980). Parameter space design of robust control systems. *IEEE Trans. Aut. Control*, **AC-25**, 1058–1072.

Ackermann, J. E. (1985). *Sampled Data Control Systems, Analysis and Synthesis, Robust System Design*. Springer-Verlag, Berlin.

Ackermann, J. E., C. V. Hollot, R. Tempo and H. Hu (1990). On extreme point results for time-delay systems *Proc. of the American Control Conf.* San Diego, CA

Anderson, B. D. O., E. I. Jury and M. Mansour (1987). On robust Hurwitz polynomials. *IEEE Trans. on Aut. Control*, **AC-32**, 909–913.

Barmish, B. R. (1983). Invariance of the strict Hurwitz property for polynomials with perturbed coefficients. *Proc.*

IEEE Conf. on Decision and Control, San Antonio, TX; (1984) *IEEE Trans. on Aut. Control*, **AC-29**, 935–936.
Barmish, B. R. (1989). A generalization of Kharitonov's four polynomial concept for robust stability problems with linearly dependent coefficient perturbations. *IEEE Trans. on Aut. Control*, **AC-34**, 157–165.
Barmish, B. R. (1993). *New Tools for Robustness of Linear Systems*. MacMillan, New York. (To appear.)
Barmish, B. R. and C. V. Hollot (1984). Counter-example to a recent result on the stability of interval matrices by S. Bialas. *Int. J. of Control*, **39**, 1103–1104.
Barmish, B. R. and H. I. Kang (1992). Extreme point results for robust stability of interval plants: Beyond first order compensators. *Automatica*. (To appear.) (1991) *Proc. of IFAC Symposium on Design Methods for Control Systems*, Zurich, Switzerland.
Barmish, B. R. and K. H Wei (1985). Simultaneous stabilizability of single input–single output systems. *Proc. of the Int. Symposium on Mathematical Theory of Networks and Systems*, Stockholm.
Barmish, B. R., R. Tempo, C. V. Hollot and H. I. Kang (1990). An extreme point result for robust stability of a diamond of polynomials. *Proc. of the IEEE Conf. of Decision and Control*, Honolulu, Hawaii.
Barmish, B. R., C. V. Hollot, F. J Kraus and R. Tempo (1992). Extreme point results for robust stabilization of interval plants with first order compensators. *IEEE Trans. of Aut. Control*, **AC-37**, 707–714.
Bartlett, A. C. (1990). Vertex results for the steady state analysis of uncertain systems. *Proc. IEEE Conf. of Decision and Control*, Honolulu, Hawaii.
Bartlett, A. C., C. V. Hollot and L Huang (1988). Root locations of an entire polytope of polynomials: it suffices to check the edges. *Mathematics of Control, Signals and Systems*, **1**, 61–71.
Basu, S. (1989). On the multidimensional generalization of of robustness of scattering Hurwitz property of complex polynomials. *IEEE Trans. on Circuits and Systems*, **CAS-36**, 1159–1167.
Bialas, S. (1983). A necessary and sufficient condition for stability of interval matrices. *Int. J. of Control*, **37**, 717–722.
Bialas, S. (1985). A necessary and sufficient condition for the stability of convex combinations of stable polynomials or matrices. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, **33**, 473–480.
Bialas, S. and J. Garloff (1985). Convex combinations of stable polynomials. *J. of the Franklin Institute*, **319**, 373–377.
Bose, N. K. (1988). Robust multivariate scattering Hurwitz interval polynomials. *Linear Algebra and Its Applications*, **98**, 123–136.
Bose, N. K. and J. F. Delansky (1989). Boundary implications for interval positive rational functions. *IEEE Trans. on Circuits and Systems*, **CAS-36**, 454–458.
Bose, N. K. and E. Zeheb (1986). Kharitonov's theorem and stability test of multidimensional digital filters. *IEEE Proc.-G*, **133**, 187–190.
Boyd, S. P. and C. H. Barratt (1990). *Linear Controller Design: Limits of Performance*. Prentice-Hall, Englewood Cliffs, NJ.
Brewer, J. W. (1978). Kronecker products and matrix calculus in system theory. *IEEE Trans. on Circuits and Systems*, **CAS-25**, 772–781.
Chapellat, H. and S. P. Bhattacharyya (1989). A generalization of Kharitonov's theorem: robust stability of interval plants. *IEEE Trans. on Aut. Control*, **AC-34**, 306–311.
Chapellat, H., M. Dahleh and S. P. Bhattacharyya (1990). Robust stability under structured and unstructured perturbations. *IEEE Trans on Aut. Control*, **AC-35**, 1100–1108.
Chapellat, H., M. Dahleh and S. P. Bhattacharyya (1991). On robust nonlinear stability of interval control systems. *IEEE Trans. on Aut. Control*, **AC-36**, 59–67.
Chen, J. M. K. H. Fan and C. Nett (1992). The structured singular value and stability of uncertain polynomials: a missing link. *Proc. of the American Control Conf.*, Chicago, IL.
Cieslik, J. (1987). On possibilities of the extension of Kharitonov's stability test for interval polynomials to the discrete-time case. *IEEE Trans. on Aut. Control*, **AC-32**, 237–238.
Cobb, J. D. and C. L. DeMarco (1989). The minimal dimension of stable faces to guarantee stability of a matrix polytope. *IEEE Trans. of Aut. Control*, **AC-34**, 990–992.
Coxson, G. E. and C. L. DeMarco (1992). Computing the structured real singular value is NP-hard. Technical Report ECE-93-4, Department of Electrical and Computer Engineering, University of Wisconsin–Madison.
Dahleh, M. A., A. Tesi and A. Vicino (1991). On the robust Popov criterion for interval L'ure systems. Technical Report, RT 24/91, Dipartimento di Sistemi e Informatica, Universita di Firenze.
Dasgupta, S. (1987). A Kharitonov like theorem for system under nonlinear parameter feedback. *Proc. of the IEEE Conf. on Decision and Control*, Los Angeles, CA.
Dasgupta, S. (1988). Kharitonov's theorem revisited. *Systems and Control Letters*, **11**, 381–384.
Dasgupta, S. and A. S. Bhagwat (1987). Conditions for designing strictly positive real transfer functions for adaptive output error identification. *IEEE Trans. on Circuits and Systems*, **CAS-34**, 731–736.
Dasgupta, S., P. J. Parker, B. D. O. Anderson, F. J. Kraus and M. Mansour (1991). Frequency domain conditions for robust stability verification of linear and nonlinear dynamical systems. *IEEE Trans. on Circuits and Systems*, **CAS-38**, 389–397.
de Gaston, R. R. E. and M. G. Safonov (1988). Exact calculation of the multiloop stability margin. *IEEE Trans. on Aut. Control*, **AC-33**, 156–171.
Demmel, J. (1988). On structured singular values. *Proc. of the IEEE Conf. on Decision and Control*, Austin, TX.
Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proc.*, **129**, 242–251.
El Ghaoui, L. (1990). *Robustnesss of linear systems*. Ph.D. Dissertation, Department of Aeronautics and Astronautics, Stanford University.
Foo Y. K. and Y. C. Soh (1989). Root clustering of interval polynomials in the left-sector. *Systems and Control Letters*, **13**, 239–245.
Frazer, R. A. and W. J. Duncan (1929). On the criteria for the stability of small motions. *Proc. of the Royal Society*, **124**, 642–654.
Fu, M. (1987). Comments on 'a necessary and sufficient condition for the positive-definiteness of interval symmetric matrices'. *Int. J. of Control*, **46**, 1485.
Fu, M. (1991). A class of weak Kharitonov regions for robust stability of linear uncertain systems. *IEEE Trans. on Aut. Control*, **AC-36**, 975–978.
Fu, M. and B. R. Barmish (1988). Maximal unidirectional perturbation bounds for stability of polynomials and matrices. *Systems and Control Letters*, **11**, 173–179.
Gantmacher, F. R. (1959). *The Theory of Matrices*. Chelsea, New York.
Garofalo, F., G Celentano and L. Glielmo (1992). Stability robustness of interval matrices via Lyapunov quadratic forms. *IEEE Trans. on Aut. Control*. (To appear.)
Ghosh, B. K (1985). Some new results on the simultaneous stabilizability of a family of single input, single output systems. *Systems and Control Letters*, **6**, 39–45.
Hara, S., T. Kimura and R. Kondo (1991). H_∞ control system design by a parameter space approach. *Proc. of the Symposium on the Mathematical Theory of Networks and Systems*, Kobe, Japan.
Hirsh, M. A. (1992). Unification of discrete and continuous time stability criteria based on Kharitonov's theorem for uncertain systems. (To appear.)
Hollot, C. V. (1989). Kharitonov-like results in the space of Markov parameters. *IEEE Trans. on Aut. Control*, **AC-34**, 536–538.
Hollot, C. V. and A. C. Bartlett (1986). Some discrete-time counterparts to Kharitonov's stability criteria for uncertain systems. *IEEE Trans. on Aut. Control*, **AC-31**, 355–356.

Hollot, C. V. and R. Tempo (1991). On the Nyquist envelope of an interval plant family. *Proc. of the American Control Conf.*, Boston, MA.

Hollot, C. V. and F. Yang (1990). Robust stabilization of interval plants using lead or lag compensators. *Systems and Control Letters*, **14**, 9–12.

Holohan, A. M. and M. G. Safonov (1989). On computing the MIMO real structured stability margin. *Proc. of the Conf. on Decision and Control*, Tampa, FL.

Horisberger, H. P. and P. R. Belanger (1976). Regulators for linear time invariant plants with uncertain parameters. *IEEE Trans. on Aut. Control*, **AC-21**, 705–708.

Horowitz, I. M. (1963). *Synthesis of Feedback Systems*. Academic Press, New York.

Hu, H. Z., C. V. Hollot, R. Tempo and J. E. Ackermann (1991). Absence of extreme point results in sampled data control systems. *Proc. of the European Control Conf.*, Grenoble, France.

Kang, H. I. (1992). *Extreme point results for robustness of control systems*. Ph.D. dissertation, in preparation, Department of Electrical and Computer Engineering, University of Wisconsin-Madison.

Karl, W. C., J. P. Greschak and G. C. Verghese (1984). Comments on 'a necessary and sufficient condition for the stability of interval matrices'. *Int. J. of Control*, **39**, 849–851.

Kharitonov, V. L. (1978a). Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differentsial'nye Uravneniya*, **14**, 2086–2088.

Kharitonov, V. L. (1978b). On a generalization of a stability criterion. *Izvestiya Akademii Nauk Kazakhskogo SSR Seriya Fizika Materiay*, **1**, 53–57.

Kharitonov, V. L. (1979). The Routh–Hurwitz problem for families of polynomials and quasipolynomials. *Izvestiia Akademii nauk Kazakhskoi SSR, Seria fizikomatematicheskia*, **26**, 69–79.

Khargonekar, P. P. and A. Tannenbaum (1985). Non-Euclidean metrics and the robust stabilization of systems with parameter uncertainty. *IEEE Trans. on Aut. Control*, **AC-30**, 1005–1013.

Kim, K. D. and N. K. Bose (1988). Invariance of the strict Hurwitz property for bivariate polynomials under coefficient perturbations. *IEEE Trans. on Aut. Control*, **AC-33**, 1172–1174.

Kim, K. D. and N. K. Bose (1990). Vertex implications of stability for a class of delay-differential integral system. *IEEE Trans. on Circuits and Systems*, **CAS-37**, 969–972

Kimura, T. and S. Hara (1991). A robust control system design by a parameter space approach based on sign definite condition. *Proc. of Korean Aut. Control Conf.*, Seoul, Korea.

Kraus, F. J., B. D. O. Anderson, E. I. Jury and M. Mansour (1988a). On robustness of low order Schur polynomials. *IEEE Trans. on Circuits and Systems*, **CAS-35**, 570–577.

Kraus, F. J., B. D. O. Anderson and M. Mansour (1988b). Robust Schur polynomial stability and Kharitonov's theorem. *Int. J. of Control*, **47**, 1213–1225.

Kwakernaak, H. (1982). A condition for robust stabilizability. *Systems and Control Letters*, **2**, 1–5.

Leitmann, G. (1979). Guaranteed asymptotic stability for some linear sytems with bounded uncertainties. *J. of Dynamic Systems, Measurement and Control*, **101**, 212–216.

Mansour, M. (1989). Simplified sufficient conditions for robust stability of interval matrices. *Int. J. of Control*, **50**, 433–444.

Mansour, M., F. J. Kraus and B. D. O. Anderson (1988). Strong Kharitonov theorem for discrete systems. *Proc. of the IEEE Conf. on Decision and Control*, Austin, TX.

Marden, M. (1966). *Geometry of Polynomials*. American Mathematical Society, Providence, RI.

Mayer, G. (1984). On the convergence of powers of interval matrices. *Linear Algebra and Its Applications*, **58**, 201–216.

Minnichelli, R. J., J. J. Anagnost and C. A. Desoer (1989). An elementary proof of Kharitonov's theorem with extensions. *IEEE Trans. on Aut. Control*, **AC-34**, 995–998.

Mori, T. and S. Barnett (1988). On stability tests for some classes of dynamical systems with perturbed coefficients. *IMA J. of Mathematical Control and Information*, **5**, 117–123.

Mori, T. and H. Kokame (1987). Convergence property of interval matrices and interval polynomials. *Int. J. of Control*, **45**, 481–484.

Panier, E., M. K. H. Fan and A. Tits (1989). On the robust stability of polynomials with no cross-coupling between the perturbations in the coefficients of even and odd powers. *Systems and Control Letters*, **12**, 291–299.

Perez, F., D. Decampo and C. Abdallah (1992). New extreme-point robust stability results for discrete-time polynomials. *Proc. of the American Control Conf.*, Chicago, IL.

Petersen, I. R. (1989). A class of stability regions for which a Kharitonov-like theorem holds. *IEEE Trans. on Aut. Control*, **AC-34**, 1111–1115.

Petersen, I. R. (1990). A new extension to Kharitonov's theorem. *IEEE Trans, on Aut. Control*, **AC-35**, 825–828.

Rantzer, A. (1990). Hurwitz testing sets for parallel polytopes of polynomials. *Systems and Control Letters*, **15**, 99–104.

Rantzer, A. (1991). *Parametric uncertainty and feedback complexity in control systems*. Ph.D. Dissertation, Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, Sweden.

Rantzer, A. (1992a). Stability conditions for polytopes of polynomials. *IEEE Trans. on Aut. Control*, **AC-37**, 79–89.

Rantzer, A. (1992b). Kharitonov's weak theorem holds if and only if the stability region and its reciprocal are convex. *Int. J. of Nonlinear and Robust Control*. (To appear.)

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

Rohn, J. (1989) Systems of linear interval equations. *Linear Algebra and Applications*, **126**, 39–78.

Rohn, J. and S. Poljak (1992). Checking robust nonsingularity is NP-hard. *Systems and Control Letters*. (To appear.)

Saeki, M. (1986). A method of robust stability analysis with highly structured uncertainties. *IEEE Trans. on Aut Control*, **AC-31**, 935–940.

Safonov, M. G. (1982). Stability margins of diagonally perturbed multivariable feedback systems. *IEE Proceedings*, **129**, 251–256.

Shafai, B., K. Perev, D. Cowley and Y. Chehab (1991). A necessary and sufficient condition for the stability of nonnegative interval discrete systems. *IEEE Trans. on Aut. Control*, **AC-36**, 742–746.

Shi, Y Q. (1991). Robust (strictly) positive interval rational functions. *IEEE Trans. on Circuits and Systems*, **CAS-38** 552–554.

Shi, Z. and W. B. Gao (1986). A necessary and sufficient condition for the positive-definiteness of interval symmetric matrices. *Int. J. of Control*, **42**, 325–328.

Sideris, A. and B. R. Barmish (1989). An edge theorem for polytopes of polynomials which can drop in degree. *Systems and Control Letters*, **13**, 233–238.

Sideris, A. and R. S. Sanchez Pena (1989). Fast computation of the multivariable stability margin for real interrelated uncertain parameters. *IEEE Trans. on Aut. Control*, **AC-34**, 1272–1276.

Siljak, D. D. (1969). *Nonlinear Systems*. Wiley, New York.

Soh, C. B. (1990). Invariance of the aperiodic property for polynomials with perturbed coefficients. *IEEE Trans. on Aut. Control*, **AC-35**, 616–618.

Soh, C. B. (1991). Robust stability of discrete-time systems using delta operators. *IEEE Trans. on Aut. Control*, **AC-36**, 377–380.

Soh, C. B. (1992). On damping ratio of polynomials with perturbed coefficients. *IEEE Trans. on Aut. Control*, **AC-37**, 277–279.

Soh, C. B. and C. S. Berger (1988a). Damping margins of polynomials with perturbed coefficients. *IEEE Trans. on Aut. Control*, **AC-33**, 509–511.

Soh, C. B. and C. S. Berger (1988b). Damping ratio of polynomials with perturbed coefficients. *IEEE Trans. on Aut. Control*, **AC-33**, 1180–1182

Soh, Y. C. and Y. K. Foo (1990). Generalization of strong Kharitonov theorems to the left sector. *IEEE Trans. on Aut. Control*, **AC-35**, 1378–1382.

Soh, Y. C. and Y. K. Foo (1991). Kharitonov regions: it suffices to check a subset of vertex polynomials. *IEEE Trans. on Aut. Control*, **AC-36**, 1102–1105.

Tempo, R. (1989). A simple test for Schur stability of a diamond of complex polynomials. *Proc of the Conf. on Decision and Control*, Tampa, FL.

Tempo, R. (1990). A dual result to Kharitonov's theorem. *IEEE Trans. on Aut. Control*, **AC-35**, 195–197.

Vaidyanathan, P. P. (1990). Derivation of new and existing discrete-time Kharitonov theorems based on discrete-time reactances. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **ASSP-38**, 277–285.

Vidyasagar, M. and N. Viswanadham (1982). Algebraic design techniques for reliable stabilization. *IEEE Trans. on Aut. Control*, **AC-27**, 1085–1095.

Vicino, A. and A. Tesi (1991). Strict positive realness: new results for interval plants plus controller families. *Proc. of the IEEE Conf. on Decision and Control*, Brighton, U.K.

Wei, K. H. and R. K Yedavalli (1987). Invariance of a strict Hurwitz property for uncertain polynomials with dependent coefficients. *IEEE Trans. on Aut. Control*, **AC-32**, 907–909.

Zadeh, L. A. and C A. Desoer (1963). *Linear System Theory—A State-space Approach*. McGraw-Hill, New York.

Zong, S. Q. (1990). Robust stability analysis of control systems with structured parametric uncertainties and an application to a robot control system. Ph.D. Dissertation, Department of Physics and Electrical Engineering, University of Bremen, Germany.

Automatica, Vol. 29, No. 1, pp. 37–56, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00
Pergamon Press Ltd

Controller Design for Plants with Structured Uncertainty*

MUNTHER A. DAHLEH†‡ and MUSTAFA H. KHAMMASH§||

A controller design methodology based on linear programming is suggested to handle time-domain specifications in the presence of structured plant uncertainty.

Key Words—Optimal control; linear programming; robustness; time-varying systems.

Abstract—This paper addresses the problem of designing feedback controllers to achieve good performance in the presence of structured plant uncertainty and bounded but unknown disturbances. A general formulation for the performance robustness problem is presented and exact computable conditions are furnished. These conditions are then utilized for synthesizing robust controllers which involves solving ℓ_1 optimization problems. These solutions are computed using the duality theory of Lagrange multipliers. Approximations and computational issues are discussed.

1. INTRODUCTION

THE OBJECTIVE OF ROBUST CONTROL is to provide in a quantitative way the fundamental limitations and capabilities of controller design in order to achieve good performance requirements in the presence of uncertainty. Even though a real system is not uncertain, it is desirable to think of it as such to reflect our imprecise or partial knowledge of its dynamics. On the other hand, uncertainty in the noise and disturbances can be cast under "real uncertainties", as it is practically impossible to provide exact models for such inputs.

Many of the design specifications tend to be concerned with amplitudes of signals. For instance, tracking, disturbance rejection, actuator authority, all result in specifications concerning the maximum amplitudes of signals. On the other hand, disturbances and noise are usually persistent, bounded, otherwise unknown. This environment motivates a peak-to-peak kind of specifications, which is the theme of the ℓ_1 theory.

In this paper, a general framework for designing controllers that achieve robust peak-to-peak performance in the presence of plant perturbations is presented. First, computable necessary and sufficient conditions for performance robustness are presented. The connections between these conditions and spectral properties of positive matrices are highlighted and utilized to simplify the computations. These conditions are in turn used for controller synthesis. The synthesis procedure followed will involve iterative solutions of ℓ_1 minimization problems, the solution of which is obtained by using the duality theory of Lagrange multipliers.

The ℓ_1 problem, formulated in Vidyasagar (1986), was solved in Dahleh and Pearson (1987, 1988a). The theory was further developed in Deodhare and Vidyasasgar, 1990, Diaz-Bobillo and Dahleh (1993, 1992), Mendlovitz (1989) and Staffans (1990, 1991). The robust stabilization problem in the presence of ℓ_∞-stable perturbations was first analysed in Dahleh and Ohta (1988) in the case of unstructured perturbations. In Khammash and Pearson (1990) a performance objective was added to the robust stability requirement in the unstructured perturbations case and conditions were provided for robust performance and stability. This led the way to the development of exact necessary and sufficient conditions for robust performance in the presence of structured perturbations (Khammash and Pearson (1991a, b, 1992)). Most of the above results have continuous-time analogs.

* Received 1 November 1991; received in final form 4 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor B. A. Francis.

† This research was supported by Wright Patterson AFB under grant F33615-90-C-3608, C.S. Draper Laboratory under grant DL-H-441636 and NSF under grant 9157306-ECS.

‡ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

§ This research is partially supported by NSF grant ECS-9110764.

|| Electrical Engineering and Computer Engineering Department, Iowa State University, Ames, IA 50011, U.S.A.

This paper is intended to be of tutorial value. In addition to surveying some of the important relevant results, new results are also included. Presented in this paper is a unified framework for designing robust controllers in the presence of structured uncertainty. Non-conservative conditions to guarantee robust performance are developed directly in terms of the spectral radius of certain matrices capturing the structure of the perturbations. Exact relations between these conditions and linear matrix inequality conditions are then established. In addition, the use of linear programming in synthesizing robust controllers is highlighted through the application of the theory of Lagrange multipliers. Through this simple formulation, problems that admit a finite-dimensional equivalence become quite transparent. For the rest of the problems, the theory proves to be quite instrumental in providing upper and lower approximations of the exact problem.

This paper puts together all of the above development in a way that makes the theory readily usable for design. In general, details that appeared elsewhere will not be presented, however, simple and intuitive proofs of the main ideas will be. Similarities and contrasts between this theory and the μ formalism will also be highlighted.

2. PRELIMINARIES

First, some notation regarding standard concepts for input/output systems. For more details, consult Desoer and Vidyasagar (1975) and Willems (1971) and the references therein.

- ℓ_∞^N denotes the set of all sequences $f = \{f_0, f_1, f_2, \ldots\}$, $f_k \in \mathbb{R}^N$, such that

$$\|f\|_{\ell_\infty} = \sup_k |f(k)|_\infty < \infty,$$

  where $|f(k)|_\infty$ is the standard ℓ_∞ norm on vectors. Also, $\ell_{\infty,e}$ denotes the extended space of all sequences in $\mathbb{R}^N$ and $\ell_{\infty,e}\backslash\ell_\infty$ denotes the set difference.
- ℓ_p, $p \in [1, \infty)$, denotes the set of all sequences such that

$$\|f\|_{\ell_p} = \left(\sum_k |f(k)|_p^p\right)^{1/p} < \infty.$$

- c_0 denotes the subspace of ℓ_∞ of sequences converging to zero.
- S denotes the backward shift operator (unit time delay).
- P_k denotes the kth-truncation operator on $\ell_{\infty,e}$:

$$P_k : \{f_0, f_1, f_2, \ldots\} \to \{f_0, \ldots, f_k, 0, \ldots\}.$$

A nonlinear operator $H : \ell_{\infty,e} \to \ell_{\infty,e}$ is causal if

$$P_k H = P_k H P_k, \quad \forall k = 0, 1, 2, \ldots,$$

strictly causal if

$$P_k H = P_k H P_{k-1}, \quad \forall k = 0, 1, 2, \ldots,$$

time-invariant if it commutes with the shift operator ($HS = SH$), and ℓ_p stable if

$$\|H\| = \sup_k \sup_{\substack{f \in \ell_{p,e} \\ P_k f \neq 0}} \frac{\|P_k H f\|_{\ell_p}}{\|P_k f\|_{\ell_p}} < \infty.$$

The quantity $\|H\|$ is called the induced operator norm over ℓ_p.

$\mathcal{L}_{TV}$ denotes the set of all linear causal ℓ_∞-stable operators. This space is characterized by infinite block lower triangular matrices of the form

$$\begin{pmatrix} H_{00} & & 0 \\ H_{10} & H_{11} & \\ \vdots & \vdots & \ddots \end{pmatrix}$$

where H_{ij} is a $p \times q$ matrix. This infinite matrix representation of H acts on elements of ℓ_∞^q by multiplication, i.e. if $u \in \ell_\infty^q$, then $y := Hu \in \ell_\infty^p$ where $y(k) = \sum_{j=0}^{k} H_{kj} u(j) \in \mathbb{R}^p$. The induced norm of such an operator is given by:

$$\|H\|_{\mathcal{L}_{TV}} = \sup_i |(H_{i1} \cdots H_{ii})|_1,$$

where $|A|_1 = \max_i \sum_j |a_{ij}|$.

- $\mathcal{L}_{TI}$ denotes the set of all $H \in \mathcal{L}_{TV}$ which are time-invariant. It is well known that $\mathcal{L}_{TI}$ is isomorphic to ℓ_1 and the matrix representation of the operator has a Toeplitz structure. Every element in $\mathcal{L}_{TI}$ is associated with a λ-transform defined as

$$\hat{H}(\lambda) = \sum_{k=0}^{\infty} H(k)\lambda^k.$$

The collection of all such transforms is usually denoted by **A**, which will be equipped with the same norm as the ℓ_1 norm.

Throughout this paper, systems are thought of as operators. So the composition of two operators G, H is denoted as GH. If both are time-invariant then $GH \in \ell_1$ (or $\mathcal{L}_{TI}$), and the induced norm is denoted by $\|GH\|_1$. When the λ-transform is referred to specifically, we use the notation $\hat{H}$ for the transform of H. Also, all operator spaces are matrix-valued functions whose dimensions will be suppressed in general whenever understood from the context.

Let X be a normed linear space. The space of all bounded linear functionals on X is denoted X^*, equiped with the natural induced norm; X^*

is always complete. It is convenient to put on X^* a weaker topology which makes $X^{**} = X$. This is the weak*-topology.

Dual of ℓ_p, $1 \le p < \infty$. The dual of ℓ_p is ℓ_q, where $(1/p) + (1/q) = 1$. The characterization is given by the following theorem.

Theorem 1. Every bounded linear functional f on ℓ_p, $1 \le p < \infty$, is representable uniquely in the form

$$f(x) = \sum_{i=0}^{\infty} x_i y_i,$$

where $y = (y_i)$ is an element in ℓ_q. Furthermore, every element of ℓ_q defines a member of ℓ_p^* in this way and

$$\|f\| = \|y\|_q.$$

The above definitions are extended for vector-valued sequences and matrix-valued sequences in the obvious way.

In this paper, we will give a solution to the ℓ_1 synthesis problem by using the theory of Lagrange multipliers. Many people are quite familiar with this theory for finite-dimensional optimization problems, and in the sequel, we will review the basic duality theorem for infinite-dimensional problems. For a more thorough treatment, see Luenberger (1969).

Let X be a vector space. A *convex cone* P is a convex set such that if $x \in P$ then $\alpha x \in P$ for all real $\alpha \ge 0$. Given such P, it is possible to define an ordering relation on X as follows: $x \ge y$ if and only if $x - y \in P$. Then it is natural to define a dual cone P^* (with an abuse of notation) inside X^* in the following way:

$$P^* = \{x^* \in X^* \mid \langle x, x^* \rangle \ge 0 \, \forall x \in P\}.$$

This in turn defines an ordering relation on X^*.

Let f be a convex function from X to $\mathbb{R}$ and G a convex map from X to another normed space Z. Also, let Ω be a convex subset of X. Assume that there exists $x_1 \in X$ such that $G(x_1) < 0$ (the inequality with respect to some cone in Z). This is generally known as the regularity assumption. Define the minimization problem:

$$\mu_0 = \inf f(x) \quad \text{subject to} \quad x \in \Omega, \quad G(x) \le 0.$$

The Lagrange multiplier theory basically says that this constrained optimization problem can be transformed to an unconstrained problem over $x \in \Omega$. Precisely, there exists an element $z_0^* \ge 0$ in Z^* (with respect to the dual cone) such that

$$\mu_0 = \inf_{x \in \Omega} \{f(x) + \langle G(x), z_0^* \rangle\}.$$

The element z_0^* is precisely the Lagrange multiplier. Equivalently,

$$\mu_0 = \sup_{z^* \ge 0} \inf_{x \in \Omega} \{f(x) + \langle G(x), z^* \rangle\}.$$

In the case where the infimization problem contains equality constraints, we will replace them by two inequality constraints. Care should be taken in this case since the assumption that the constraint set has an interior point will be violated; however under mild assumptions, if the equality constraints are given in terms of linear operators, the result will still hold without the regularity conditions.

3. WHY THE ℓ_∞ SIGNAL NORM?

In many real-world applications, output disturbance and/or noise is persistent, i.e. continues acting on the system as long as the system is in operation. This implies that such inputs have infinite energy, and thus one cannot model them as "bounded-energy signals". Nevertheless, one can get a good estimate on the maximum amplitude of such inputs. Examples where bounded disturbances arise in practical situations are abundant. Wind gusts facing an aircraft in flight can be viewed as bounded disturbances. Without a correcting control action, such disturbances will cause the aircraft to deviate from its set path. An automobile driven over an unpaved road experiences disturbances due to the irregularity of the course. Such disturbances, although persistent, are clearly bounded in magnitude. In process control, level measurements of a boiling liquid are corrupted by bounded disturbances due to the constant level fluctuations of the liquid. Because such disturbances are so frequent, a mathematical model describing them is essential. The ℓ_∞ norm is clearly the most natural choice for measuring the size of such disturbances. In general, we will assume that the disturbance is the output of a linear time-invariant (LTI) filter subjected to signals of magnitude less than or equal to one, i.e.

$$d = Ww, \quad \|w\|_\infty \le 1.$$

Not only is the ℓ_∞ norm useful for measuring input signal size, but it can also be very useful as a measure for the size of output signals. For example, in many applications it is crucial that the tracking error never exceeds a certain level at *any* time. While this requirement cannot be captured by using the ℓ_2 norm, it can be stated explicitly as a condition on the ℓ_∞ norm of the error signal. Another situation when the ℓ_∞ norm is useful is when the plant, or any other device in the control loop, has a maximum input rating which should not be exceeded. This translates

directly to a requirement on the ℓ_∞ norm of that input. An example of such a requirement appears in the next section. In addition, the ℓ_∞ norm plays an important role in designing controllers for nonlinear systems. Since most of the nonlinear controller designs are based on linearization, the linear model gives a faithful representation of the system only if the states *remain close to the equilibrium point*, a requirement captured directly in terms of the ℓ_∞ norm.

4. THE ℓ_1 NORM

While the ℓ_∞ norm is used as a measure of signal size, the ℓ_1 norm is used to measure a system's amplification of ℓ_∞ input signals. Let T be an LTI system given by

$$z(t) = (Tw)(t) = \sum_{k=0}^{t} T(k)w(t-k).$$

The inputs and outputs of the system are measured by their maximum amplitude over all time, otherwise known as the ℓ_∞ norm, i.e.

$$\|w\|_\infty = \max_j \sup_k |w_j(k)|.$$

The ℓ_1 norm of the system T is precisely equal to the maximum amplification the system exerts on bounded inputs. This measure defined on the system T is known as the induced operator norm and is mathematically defined as

$$\|T\| = \sup_{\|w\|_\infty < 1} \|Tw\|_\infty = \|T\|_1,$$

where $\|T\|_1$ is the ℓ_1 norm of the pulse response and is given by

$$\|T\|_1 = \max_i \sum_j \sum_k |t_{ij}(k)|.$$

A system is said to be ℓ_∞-stable if it has a bounded ℓ_1 norm, and the space of all such systems will be denoted by ℓ_1. From this definition, it is clear that the system attenuates inputs if its ℓ_1 norm is strictly less than unity.

In the case where the inputs and outputs of the linear system are measured by the ℓ_2 norm, then the gain of the system is given by the H_∞ norm and is given by Doyle *et al.* (1989), Francis (1987), Vidyasagar (1985) and Zames (1981)

$$\|\hat{T}\|_\infty = \sup_{0 \le \theta \le 2\pi} \sigma_{\max}(\hat{T}(e^{i\theta})).$$

The two induced norms are related by Boyd and Doyle (1987)

$$\|\hat{T}\|_\infty \le C_1 \|T\|_1 \le C_2(N) \|\hat{T}\|_\infty,$$

where C_1 is a constant depending only on the dimension of the matrix T, and C_2 is a linear function of the McMillan degree N of $\hat{T}$. In other words, every system inside ℓ_1 is also inside H_∞, but the converse is not true. This means that there exist ℓ_2 stable LTI systems that are not ℓ_∞ stable; an example is the function with the λ-transform given by Boyd and Doyle (1987);

$$\hat{T}(\lambda) = e^{1/(1-\lambda)}.$$

Thus, for LTI systems, *minimizing the ℓ_1 norm* guarantees that the H_∞ norm is bounded. This means that this system will have good ℓ_2-disturbance rejection properties as well as ℓ_∞-disturbance rejection properties. Also, the ℓ_1 norm is more closely allied with BIBO stability notions and hence is quite desirable to work with. The disadvantage in working with the ℓ_1 norm is the fact that it is a Banach space of operators operating on a Banach space, not a Hilbert space itself. Many of the standard tools are not usable; however, this paper will present new techniques for handling problems of this kind.

5. PROTOTYPE PROBLEMS

In this section we demonstrate the advantages of using the ℓ_∞ signal norm by presenting a few prototype problems. For each problem, certain control objectives related to the ℓ_∞ norm are to be met. These problems demonstrate the advantages of using the ℓ_∞ signal norm as a means of capturing time-domain specifications in an uncertain environment. Later on, it will be shown how all such problems can be treated in a unified manner under a single framework. We shall then develop mathematical techniques for obtaining solutions for all problems which fit within that framework.

5.1. *Disturbance rejection problem*

Consider the system in Fig. 1. Here P_0 is a plant and K a controller, both LTI. The system is subjected to bounded disturbances which are reflected at the plant output. As mentioned earlier, these disturbances are assumed to be the output of a time-invariant filter W_1 reflecting the frequency content of such disturbances. The

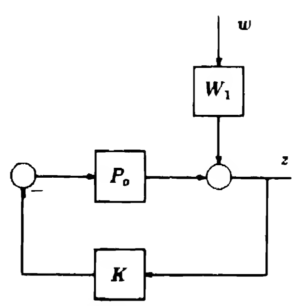


FIG. 1. Disturbance rejection problem

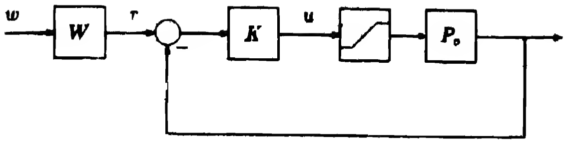


FIG 2 Command following with input saturation.

control objective in this case will be to find a controller K which satisfies the following:

(1) K internally stabilizes the feedback system.
(2) The effect of the disturbances at the plant output is minimized, i.e. K minimizes $\sup_{\|w\|_\infty \le 1} \|z\|_\infty$.

5.2. *Command following in the presence of input saturation*

The command following problem is equivalent to the disturbance rejection problem. Consider the system in Fig. 2. The plant, P, suffers from saturation nonlinearities at its input. Therefore, it can be viewed as having two components: a saturation component, Sat (.), and an LTI component, P_0. The saturation component is defined as follows:

$$\text{Sat}(u) = \begin{cases} u & |u| < U_{\max} \\ U_m & |u| \ge U_{\max} \end{cases}$$

As a result the plant is described as $P = P_0\,\text{Sat}(.)$. Because of the presence of the saturation, the plant input, u, must not be allowed to exceed $U_{\max}$. This requirement can be captured in a natural way using the ℓ_∞ norm of u. In other words, u must satisfy $\|u\|_\infty \le U_{\max}$.

The command, r, is to be followed at the plant output. It is not fixed but rather can be any command in the set

$$\{r = Ww : \|w\|_\infty \le 1\},$$

where W reflects the frequency content of the desired commands and is typically a low pass filter.

The control objectives can now be stated more precisely. It is desired to find a controller K such that:

(1) K internally stabilizes the system.
(2) $\|u\|_\infty \le U_{\max}$.
(3) y follows r uniformly in time to within a maximum error level of $\gamma > 0$, i.e. $\|y - r\|_\infty \le \gamma$.

Here, by internal stability we mean local internal stability. As long as the external signals do not cause u to exceed $U_{\max}$ the system will be operating in the linear region and internally stability of the linear system implies that all signals in the loop will be bounded for bounded inputs.

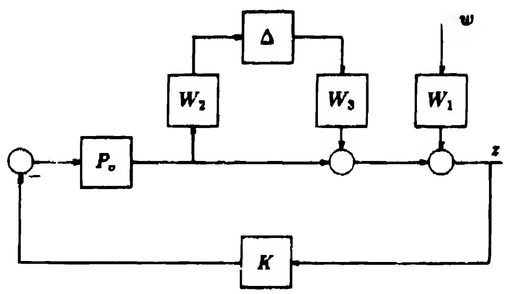


FIG 3 Robust disturbance rejection problem.

5.3. *Robust disturbance rejection*

In the previous two problems, the plant was assumed to be known exactly. This is rarely the case due to unmodelled dynamics, parameter variations, etc. When the controller designed for a nominal plant model is implemented on the real system, there are no guarantees on the resulting performance of the system. Even requirements as basic as stability may not be met. The deviation from the expected behavior of the system clearly depend on the accuracy of the model. Since modelling uncertainty is inevitable, it is imperative to include stability and performance robustness to model uncertainty as a design objective. We now take a second look at the disturbance rejection problem discussed earlier. Instead of considering a single nominal time-invariant plant, P_0, we shall instead consider a collection of plants. The class of plants considered is taken to be

$$\Pi := \left\{ P = (I + W_3 \Delta W_2) P_0 \; : \Delta \text{ is causal and } \|\Delta\| := \sup_{u \ne 0} \frac{\|\Delta u\|_\infty}{\|u\|_\infty} \le 1 \right\}$$

where W_1 and W_2 are time-invariant weighting functions (see Fig. 3). In this definition, the plant perturbation, Δ, may be time-varying and/or nonlinear. Any plant belonging to this plant class is said to be admissible. Note that when $\Delta = 0$, we recover the nominal LTI plant. Consequently, the collection of admissible plants, Π, may be viewed as a ball of plants centered around the nominal time-invariant plant model. If a system property, such as stability, holds for all admissible plants it is said to be *robust*. We now add to our original disturbance rejection problem a new objective: robustness. In other words, the controller K is now required to perform the following tasks:

(1) K internally stabilizes all admissible plants, i.e. all plants in the class Π.
(2) K minimizes the effect of the disturbance w on the magnitude of the output for the worst

possible admissible plant, i.e. K minimizes $\sup_{P\in\Pi}\sup_{\|w\|_\infty\le 1}\|y\|_\infty$.

5.4. *Robustness in the presence of coprime factor perturbations*

Another approach to the representation of plant uncertainty is through coprime factor perturbations (Dahleh (1992); Glover and McFarlane (1989)). Let $P_0 = NM^{-1}$ be a coprime factorization of the nominal plant. The graph of the plant P_0 over the space ℓ_∞ is defined as the image of the space ℓ_∞ under the map G_{P_0} where

$$G_{P_0}: \ell_\infty \mapsto \ell_\infty \times \ell_\infty,$$

$$G_{P_0}u = \begin{bmatrix} Mu \\ Nu \end{bmatrix}.$$

The class of admissible plants can be defined as those plants whose graph is perturbed in the following way:

$$\Pi = \left\{P : G_P = \begin{bmatrix} M + \Delta_1 \\ N + \Delta_2 \end{bmatrix},\ \|\Delta_1\| \le 1,\ \|\Delta_2\| \le 1\right\}.$$

This plant class can be viewed as that obtained by perturbing the plant numerator and the plant denominator independently as shown in Fig. 4. The main objective in this case is to find a controller K which stabilizes all plants in the class Π.

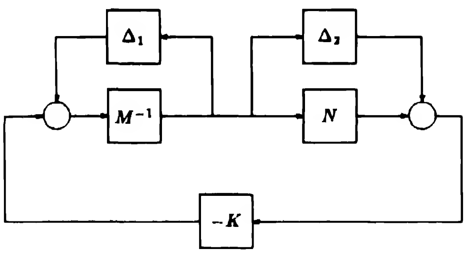


FIG. 4. Coprime factor perturbations.

5.5. *A multiobjective control problem*

In almost all practical control problems, more than one objective must be met simultaneously. Perhaps one of the most attractive features of the present approach is its ability to handle multiple objectives in a natural way. As an example of a multiple objective problem consider the system in Fig. 5. In the figure the plant is subjected to multiplicative output perturbations. In addition it has a saturation nonlinearity at its input of the type discussed earlier. A command input, r, is applied while a bounded disturbance, d, is acting at the plant output. The objectives in this problem are a combination of those objectives in the first three problems discussed earlier. Aside from stabilizing all admissible plants, the controller must also ensure that the plant input, u, never exceeds its maximum, $U_{\max}$, despite the presence of the output disturbance, the command input, and the plant uncertainty. Furthermore, the tracking error in this unfriendly environment must be maintained at a minimum level for all time. These requirements on the controller are summarized as follows:

(1) K stabilizes all plants in Π.

(2) K is chosen such that $\sup_{\|w_i\|_\infty\le 1}\sup_{P\in\Pi}\|u\|_\infty \le U_{\max}$.

(3) K is chosen such that $\sup_{\|w_i\|_\infty\le 1}\sup_{P\in\Pi}\|e\|_\infty$ is minimized.

Comment.

It is possible in this formulation to include time-varying weights with which one can emphasize certain periods of the time response. The general framework and solutions presented in the sequel generalize in the presence of such weights; however, we will restrict our discussion to the time-invariant case.

6. A GENERAL FORMULATION: THE ROBUST PERFORMANCE PROBLEM

In the previous section, we have formulated sample control problems which reflect various practical control requirements. Two assumptions were embedded in the problem statement. The first of these is that the command signals and the disturbance signals do not necessarily decay in time but can instead persist over all time so long as they are bounded. This is a fairly realistic assumption and leads to the adoption of the ℓ_∞

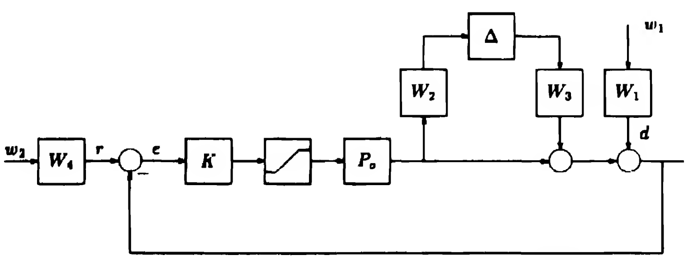


FIG. 5. Multiobjective problem.

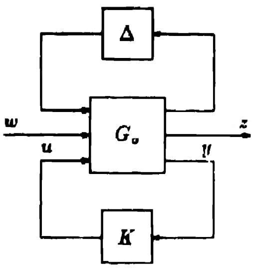


FIG. 6. The robust performance problem.

norm to measure the signal size. The second consists of requiring the regulated signals of interest to have small maximum amplitudes. Thus, once more, the ℓ_∞ signal norm is used as a measure for signal size, but this time it is the regulated output signals which are being measured. When considering that quite often the output of interest is a tracking error, plant input and/or plant output, it becomes clear that limiting the maximum value these signals can achieve is desirable if not necessary. As a means for obtaining a unifying framework for formulating and solving a wide variety of problems with ℓ_∞ signal norms and ℓ_∞ induced-norm bounded perturbations, we set up the Robust Performance Problem. All the prototype problems discussed in the previous section are special cases of this general problem. So consider the system in Fig. 6: Δ models the system uncertainty, K is the controller, and G_0 contains the remaining part of the system. It is assumed that Δ belongs to the following class:

$$\mathscr{D}(n) := \{\Delta = \operatorname{diag}(\Delta_1, \ldots, \Delta_n) : \Delta_i \text{ is causal and } \|\Delta_i\| \le 1\}.$$

Here $\|\Delta_i\|$ is the induced ℓ_∞ norm, i.e.

$$\|\Delta_i\| = \sup_{u \neq 0} \frac{\|\Delta_i u\|_\infty}{\|u\|_\infty}.$$

In the sequel, the Δ_is are assumed to be SISO for simplicity. There is no time-invariance restriction on the perturbations, and hence time-varying and/or nonlinear perturbations are allowed. The diagonal structure of the perturbations is essential for incorporating information about the location of the system uncertainty. For example actuator unmodelled dynamics are not related to sensor unmodelled dynamics or to the plant's unmodelled high-frequency dynamics, and should not be modelled by the same perturbation block. By isolating the independent sources of uncertainty, a more realistic and less conservative system model is obtained. This is the main reason for considering structured perturbations. While Δ models the uncertain part of the system, G_0 is the known part of the system with the exception of the controller, and it is a 3×3 block matrix. The actual system is an element in the upper linear fractional connection of G_0 and the admissible Δs. So, included in G_0 is the nominal plant/plants, any input and output weighting functions, and any weighting functions on the perturbations. We shall restrict the weights and the nominal plant to be LTI. As a result, G_0 is LTI. The controller K is also assumed to be LTI. The signal w denotes all exogenous inputs, including the command inputs and the disturbance inputs which are assumed to be in ℓ_∞, while z denotes the regulated outputs. Both w and z are allowed to be vector signals. From now on, we shall refer to the map taking w to z as $\mathscr{T}_{zw}$. The induced ℓ_∞ norm of $\mathscr{T}_{zw}$ is defined as follows:

$$\|\mathscr{T}_{zw}\| := \sup_{w \neq 0} \frac{\|\mathscr{T}_{zw} w\|_\infty}{\|w\|_\infty} = \sup_{w \neq 0} \frac{\|\;\|}{\|w\|}$$

We will say that the system in Fig. 6 is ℓ^∞ stable (or just internally stable) if, when injecting external ℓ^∞ signals at the inputs of Δ, G_0, and K, the effect of these signals together with $w \in \ell^\infty$ is to produce ℓ^∞ signals at any point in the loop. Furthermore, the induced norm of this mapping between the injected signals together with w and any internal signal in the loop must be bounded (see Desoer and Vidyasagar, 1975).

We are now ready to state the Robust Performance Problem.

Robust Performance Problem. Find a controller K such that:

(1) The system achieves robust stability, i.e. K internally stabilizes the system for all admissible perturbations, i.e. for all Δ in $\mathscr{D}(n)$.
(2) The system achieves robust performance, i.e. K is chosen so that

$$\sup_{\Delta \in \mathscr{D}(n)} \|\mathscr{T}_{zw}\| < 1.$$

As mentioned earlier, the prototype problems discussed can all fit in this framework. As an example, for the Disturbance Rejection Problem since the number of perturbation blocks, n, is zero, G_0 has only two inputs w and u, and two outputs z and y. As a result G_0 has the form:

$$G_0 = \begin{pmatrix} W_1 & -P_0 \\ W_1 & -P_0 \end{pmatrix}.$$

This is referred to as the nominal performance problem.

On the other hand, for the robust disturbance rejection problem n will be 1. Thus, G_0 has an additional input fed from the perturbation

output, and an additional output feeding the perturbation input. It follows that G_0 has the following structure:

$$G_0 = \begin{pmatrix} 0 & 0 & -W_2P_0 \\ W_3 & W_1 & -P_0 \\ W_3 & W_1 & -P_0 \end{pmatrix}, \text{ etc.}$$

7. ROBUSTNESS CONDITIONS

Having stated the Robust Performance Problem, we can now focus our attention on its solution. In particular, we shall develop necessary and sufficient conditions for achieving both performance robustness and stability robustness. These conditions will be used for the robustness analysis of the system at hand. In this case, the controller is assumed given and fixed and its effect on the robustness of the system is investigated. The same conditions developed for robustness analysis are used to develop techniques for the synthesis of robust controllers.

We begin by discussing the robustness analysis issue. Suppose we are given a nominal system G_0, a perturbation class $\mathscr{D}(n)$, and a controller K connected as shown in Fig. 6. We can incorporate G_0 and K together and view them as one system, M, as shown in Fig. 7. Thus M will have two inputs and two outputs. We will assume that the controller K stabilizes the nominal system G_0; otherwise robust stability and hence performance clearly will not be achieved. Consequently, M will be LTI and stable. We will say the system in Fig. 7 achieves robust stability if it is stable for all admissible perturbations. We will say that it achieves robust performance if, in addition, $\|\mathscr{T}_{zw}\| < 1$ for all admissible perturbations. We can now state the following problem whose solution is provided in the next two sections.

Robustness analysis problem. Under what conditions on M will the system in Fig. 7 achieve robust performance?

7.1. *Stability robustness vs performance robustness*

It is an interesting fact that a robust performance problem can be transformed to a robust stability problem. This has been shown in Doyle and Stein (1981) and Doyle (1982) when

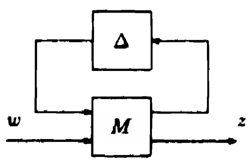


FIG. 7. Robust performance problem in M–Δ form

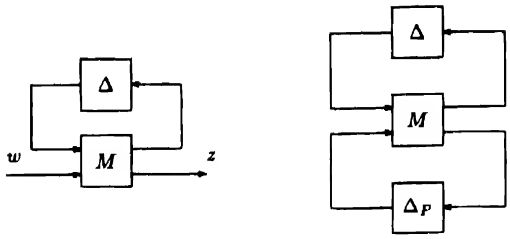


FIG. 8. Stability robustness vs performance robustness.

the perturbations are LTI with an ℓ_2 induced-norm. This remains true in our case as well, although the perturbation class and the method of proof are quite different. To elaborate further on this relationship between stability robustness and performance robustness consider the two Systems shown in Fig. 8. System I corresponds to a performance robustness problem, while System II is formed from System I by feeding z back to w through a fictitious perturbation, Δ_P, satisfying $\|\Delta_P\| \le 1$. As a result, System II has $\mathscr{D}(n+1)$ as its perturbation class. We can now ask the following question: How does the performance robustness of System I relate to the stability robustness of System II? One aspect of the relationship between the two notions of stability is fairly obvious: performance robustness of System I implies stability robustness of System II. This is quite easy to see. Since robust performance is equivalent to the norm of the map between w and z being less than one, the Small Gain Theorem can be used to establish the stability of System II for all $\|\Delta_P\| \le 1$, or equivalently to establish the robust stability of System II. Equally important, the relation between stability robustness and performance robustness holds the other way as well. In other words, stability robustness of System II implies performance robustness of System I. This direction is not as obvious as the first one. The proof follows from certain results on the robustness of time-varying systems.

7.2. *Stability robustness conditions*

Because performance robustness is equivalent to stability robustness in the sense discussed earlier, we need only discuss stability robustness. Specifically, we can consider the interconnection of a stable LTI system, M, with a structured perturbation $\Delta \in \mathscr{D}(n)$ in Fig. 9, and seek necessary and sufficient conditions for the stability robustness of the system. Since M and Δ are both stable, the internal stability of the system is equivalent to the map $I - M\Delta$ having a stable inverse, one which maps ℓ_∞ to itself with a finite gain. When the signal norm is the ℓ_2 norm

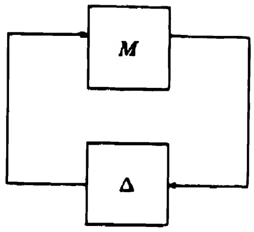


FIG. 9 Stability robustness problem.

and the perturbations are time-invariant, the conditions are provided by the Structured Singular Value, μ (Doyle, 1982). In particular, robustness is achieved iff $\sup_{0 \le \theta \le 2\pi} \mu(M(e^{j\theta})) < 1$. In our formulation, it turns out that the conditions are much simpler and easier to compute than μ. Before we can present these conditions we need to define a certain nonnegative matrix, $\hat{M}$, which depends solely on M. Recalling M has n inputs and n outputs, it can be partitioned as follows:

$$M = \begin{bmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & & \vdots \\ M_{n1} & \cdots & M_{nn} \end{bmatrix}$$

Each M_{ij} is LTI and stable, and thus $M_{ij} \in \ell_1$. Clearly $\|M_{ij}\|_1$ can be computed with arbitrary accuracy by considering finite truncations of M_{ij} as approximations. We can now define $\hat{M}$ as follows:

$$\hat{M} = \begin{bmatrix} \|M_{11}\|_1 & \cdots & \|M_{1n}\|_1 \\ \vdots & & \vdots \\ \|M_{n1}\|_1 & \cdots & \|M_{nn}\|_1 \end{bmatrix}$$

One of the most interesting aspects of the robustness problem formulated here is the role which $\hat{M}$ plays in the system robustness. This is presented in the next theorem due to Khammash and Pearson (1991a, b, 1992).

Theorem 2. The system in Fig. 9 possesses robust stability iff any one of the following holds.

(1) $\rho(\hat{M}) < 1$, where $\rho(.)$ denotes the spectral radius.
(2) $x \le \hat{M}x$ and $x \ge 0$ imply that $x = 0$, where the vector inequalities are to be interpreted componentwise.
(3) $\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 < 1$, where

$$\mathcal{R} := \{\mathrm{diag}\,(r_1, \ldots, r_n) : r_i > 0\}.$$

One of the main contributions of this theorem is that it provides simple and exact conditions for testing the system's stability robustness regardless of the number of perturbation blocks, n. While the three conditions in the theorem are equivalent, each provides a different perspective and has certain advantages over the others. For example, the spectral radius condition is in general the easiest to compute. It is particularly useful when n is large since it can be computed efficiently using power methods. Specifically, given an $\hat{M}$ which is assumed primitive (i.e. $\hat{M}^k > 0$ for some integer k) then it satisfies the following:

$$\min_i \frac{(\hat{M}^{k+1}x)_i}{(\hat{M}^k x)_i} \le \rho(\hat{M}) \le \max_i \frac{(\hat{M}^{k+1}x)_i}{(\hat{M}^k x)_i}$$

for any vector $x > 0$. Furthermore, the upper and lower bounds both converge to $\rho(\hat{M})$ as k goes to infinity. If $\hat{M}$ were not primitive, it can be perturbed slightly to become primitive.

Whereas the spectral radius test provides a yes or no answer concerning system robustness, the second test involving the Linear Matrix Inequality (LMI) is most useful for providing information about the effect of the individual entries of $\hat{M}$ on the overall robustness of the system. This is achieved by translating the LMI condition into n algebraic conditions stated explicitly in terms of the entries of $\hat{M}$. This is best demonstrated by an example. Suppose $\hat{M}$ is a 2×2 matrix corresponding to a certain robustness problem with $n = 2$. The LMI condition states that robust stability iff the system

$$x_1 \le \|M_{11}\|_1 x_1 + \|M_{12}\|_1 x_2,$$
$$x_2 \le \|M_{21}\|_1 x_1 + \|M_{22}\|_1 x_2,$$

has no solution $x = (x_1, x_2) \in [0, \infty) \times [0, \infty)\backslash\{0\}$. Among other things, this implies that $\|M_{11}\|_1 < 1$; otherwise $x = (1, 0)$ would be a solution for the two inequalities. The first inequality can be rewritten as

$$x_1 \le \frac{\|M_{12}\|_1}{1 - \|M_{11}\|} \cdot x_2.$$

When combined with the second inequality, we have that

$$x_2 \le \left(\|M_{21}\|_1 \frac{\|M_{12}\|_1}{1 - \|M_{11}\|_1} + \|M_{22}\|_1\right) x_2,$$

has no solution in $(0, \infty)$, which is equivalent to

$$\|M_{21}\|_1 \frac{\|M_{12}\|_1}{1 - \|M_{11}\|_1} + \|M_{22}\|_1 < 1.$$

This last condition, together with the condition that $\|M_{11}\|_1 < 1$, is therefore necessary for the inequality robustness conditions to hold. By retracing our steps backwards, it becomes clear that they are also sufficient. This procedure of constructing explicit norm conditions from the second robustness condition can be repeated in the same way for any n.

Finally, the third robustness condition is useful for robust controller synthesis. This will be discussed in more detail later on.

7.2.1. *Equivalence of the robustness conditions in Theorem* 2. Before we shed more light on the proof of Theorem 2, we will show that the three, apparently unrelated, conditions in the statement of the theorem are indeed equivalent. We will show that $1 \Leftrightarrow 2$ and that $1 \Leftrightarrow 3$. For simplicity, we will do this for an irreducible $\hat{M}$. So suppose that $\rho(\hat{M}) < 1$. It follows that $(I - \hat{M})^{-1}$ exists. Since $(I - \hat{M})^{-1} = I + \hat{M} + \hat{M}^2 + \cdots$, all of its entries will be positive. Now if $x \geq 0$ is such that $x \leq \hat{M}x$, or equivalently, $(I - \hat{M})x \leq 0$, then multiplying both sides by $(I - \hat{M})^{-1}$ implies that $x \leq 0$. Thus x must be zero. This is what 2 states. To show that 2 implies 1, suppose 1 does not hold, i.e., that $\rho(\hat{M}) \geq 1$. The Perron–Frobenius theory for nonnegative matrices states that $\rho(\hat{M})$ is itself an eigenvalue of $\hat{M}$. Moreover, associated with $\rho(\hat{M})$ we can find an eigenvector $x' > 0$. This implies that $\rho(\hat{M})x' = \hat{M}x'$, which in turns implies that 2 does not hold. Thus, we have demonstrated that $1 \Leftrightarrow 2$.

We now show $1 \Leftrightarrow 3$ by showing that $\rho(\hat{M}) = \inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1$. By definition,

$$\|R^{-1}MR\|_1 = \max_i \sum_j \|(R^{-1}MR)_{ij}\|_1$$

$$= \max_i \sum_{j=1}^{n} \frac{r_j}{r_i} \|M_{ij}\|_1.$$

The expression on the right is also equal to the induced norm of the matrix $R^{-1}\hat{M}R$ as a map from $(\mathbb{R}^n, \|.\|_\infty)$ to itself. Referring to this norm by $|.|_1$, we therefore have $\|R^{-1}MR\|_1 = |R^{-1}\hat{M}R|_1$. Since any matrix norm bounds from above the spectral radius of that matrix we have:

$$\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 = \inf_{R \in \mathcal{R}} |R^{-1}\hat{M}R|_1 \geq \rho(R^{-1}\hat{M}R) = \rho(\hat{M}).$$

But if we choose $R = \operatorname{diag}(r_1', \ldots, r_n')$, where $(r_1', \ldots, r_n')'$ is the positive eigenvector corresponding to the eigenvalue $\rho(\hat{M})$, the inequality becomes an equality and the equivalence between 1 and 3 is established. It is interesting to note that for the optimum scalings $R = \operatorname{diag}(r_1', \ldots, r_n')$, all the rows of $R^{-1}MR$ have the same norm. As will be demonstrated shortly, this fact is used to show why Condition 3 in the above theorem is necessary for system robustness.

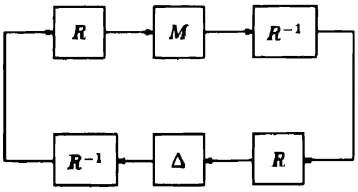


FIG. 10. Scaled system

7.2.2. *Proof of necessity and sufficiency*. When $n = 1$, the spectral radius condition in the theorem above reduces to the condition $\|M\|_1 < 1$. A simple application of the Small Gain Theorem shows that this condition is sufficient for stability. Necessity has been shown by Dahleh and Ohta (1988). For n larger than one, we now show that $\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 < 1$ is sufficient for robust stability. We do this with the aid of Fig. 10 obtained via the addition of scalings R and R^{-1}, where $R \in \mathcal{R}$. Clearly, the robustness of this system and that in Fig. 9 are equivalent in the sense that if one is robustly stable, then so is the other one. Moreover, for the system in Fig. 10. $R\Delta R^{-1}$ belongs to $\mathcal{D}(n)$ whenever Δ belongs to $\mathcal{D}(n)$, and thus $\|R\Delta R^{-1}\| \leq 1$. That being the case, the Small Gain Theorem can be invoked to conclude that $\|R^{-1}MR\|_1 < 1$ is sufficient for robust stability. This holds for any $R \in \mathcal{R}$. The least conservative sufficient condition obtainable in that manner is

$$\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 < 1.$$

It should be mentioned at this point that spectral radius type conditions have been used in the literature to give sufficient conditions for stability of large scale systems see Vidyasagar (1981) and the references therein.

We now demonstrate that $\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 < 1$ is necessary for robust stability. For simplicity, we do this for the case $n = 2$. The approach will be to show how one can construct a destabilizing perturbation $\Delta \in \mathcal{D}(2)$ whenever $\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 \geq 1$. So suppose that $\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1 \geq 1$. We have previously shown that this infimum is in fact a minimum, and it is achieved by an optimum scaling, R, obtained from the eigenvector corresponding to $\rho(\hat{M})$. It was also shown that the two rows of $R^{-1}MR$ will have equal norms. This can be expressed as follows:

$$\|(R^{-1}MR_1)\|_1 = \|(R^{-1}MR)_2\|_1 = \|R^{-1}MR\|_1 \geq 1.$$

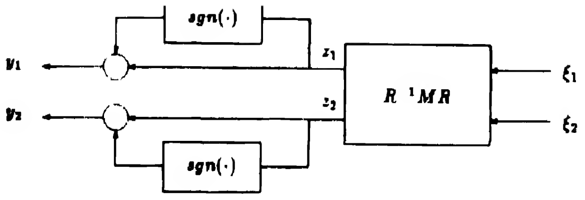


FIG. 11. Scaled system with constructed input ξ.

where $(R^{-1}MR)_i$ denotes the ith row of $R^{-1}MR$. The system $R^{-1}MR$ appears in Fig. 11 and has as its input $\xi = (\xi_1, \xi_2)$ and output $z = (z_1, z_2)$. In the figure $y = (y_1, y_2)$ consists of the output $z = (z_1, z_2)$ after a bounded signal, the output of a sign function, has been added to it. This bounded signal will be interpreted as an external signal injected for stability analysis. The strategy taken will be to construct ξ satisfying the two requirements:

(1) ξ is unbounded.
(2) ξ results in a signal y which satisfies $\|P_k\xi_i\|_1 \le \|P_k y_i\|_1$ for $i = 1, 2$, where P_k is the truncation operator which acts on sequences by preserving the first $k + 1$ terms and setting the rest to zero.

The first requirement on ξ guarantees that if an admissible perturbation Δ were to map y to ξ, it would be a destabilizing one because the bounded external signal would have produced an unbounded internal signal ξ. The second requirement, guarantees that such an admissible perturbation exists. In other words, if ξ and y satisfy the second condition, then it is possible to find Δ_i, for $i = 1, 2$, so that Δ_i is causal, has induced norm less than or equal to one, and satisfies $\Delta_i y_i = \xi_i$. If the first requirement is also met, this Δ will be a destabilizing perturbation.

For simplicity we shall assume that all M_{ij}s have finite impulse response of length, say N. The construction of ξ proceeds as follows. While maintaining $|\xi_i(k)| \le 1$ for $k = 0, \ldots, N-1$, the first N components of ξ can be constructed so as to achieve $\|(R^{-1}MR)_1\|_1$. Since $\|(R^{-1}MR)_1\|_1 \ge 1$, this implies that $\|P_{N-1}z\|_\infty \ge 1$, which in turn implies that $\|P_{N-1}y\|_\infty \ge 2$. Next, while still maintaining $|\xi_i(k)| \le 1$, we pick the next N components of ξ so as to achieve the second row norm, $\|(R^{-1}MR)_2\|_1$. As a result we have $\|P_{2N-1}z_i\|_\infty \ge 1$ which implies that $\|P_{2N-1}y_i\|_\infty \ge 2$. Note that the second requirement on ξ has been met for $k = 0, \ldots, 2N-1$. In addition, because of the way the first $2N$ terms of ξ have been constructed, we have

$$\|P_{2N-1}y_i\|_\infty \ge \|P_{2N-1}\xi_i\|_\infty + 1, \quad i = 1, 2.$$

This allows us to relax the restriction on $|\xi_i(k)|$ for $k > 2N-1$ without violating the second requirement on ξ. Specifically, we now allow $|\xi_i(k)|$ to be as large as 2 for $k = 2N, \ldots, 4N-1$. In the same way as before we can pick $\xi(k)$ for this range of k so that we satisfy

$$\|P_{4N-1}y_i\|_\infty \ge \|P_{4N-1}\xi_i\|_\infty + 1, \quad i = 1, 2,$$

which allows us to increase $|\xi_i(k)|$ by 1 for the next $2N$ components of ξ, and repeat the whole procedure again. From this construction, it is clear that when ξ is completely specified it will be unbounded and hence meets the first requirement. The second requirement is also met since all along $\xi_i(k)$ was chosen carefully so as not to become too large too soon.

It should be mentioned that the destabilizing perturbation can be taken to be linear time-varying (LTV), or it can instead be nonlinear time-invariant. So the spectral radius condition for robustness is also necessary and sufficient whenever the class of perturbation is restricted to include norm-bounded nonlinear time-invariant perturbations.

7.2.3. *Construction of the destabilizing perturbation.* In the previous section, we have claimed that given $\xi = \{\xi(i)\}_{i=0}^{\infty} \in \ell_{\infty,e}$ and $y = \{(y(i)\}_{i=0}^{\infty} \in \ell_{\infty,e}$ such that $\|P_k\xi_i\|_\infty \le \|P_k y_i\|_\infty$ $\forall k$, and for $i = 1, 2$, then there exists $\Delta = \operatorname{diag}(\Delta_1, \Delta_2)$ such that $\Delta y = \xi$ and $\|\Delta_i\| \le 1$. Such a Δ was shown to be a destabilizing perturbation. In this section, we prove this claim by explicitly constructing the perturbation Δ. It turns out that Δ_i can be either LTV or nonlinear and time-invariant. We shall construct Δ_1 to be of the former type, while Δ_2 will be of the latter type.

So suppose we are given $\xi_1 = \{\xi_1(i)\}_{i=0}^{\infty} \in \ell_{\infty,e}$ and $y_1 = \{y_1(i)\}_{i=0}^{\infty} \in \ell_{\infty,e}$ such that $\|P_k\xi_1\|_\infty \le \|P_k y_1\|_\infty$ $\forall k$. The construction of Δ_1 is trivial if $y_1 = 0$: just pick Δ_1 itself to be zero. So assume $y_1 \ne 0$. We start the construction of Δ_1 by identifying a subsequence of y_1, say $(y_1(i_1), y_1(i_2), \ldots)$ which, depending on y_1, may or may not be finite. This subsequence may be defined recursively in the following manner: Let i_1 be the smallest integer such that $y_1(i_1) \ne 0$. Given $y_1(i_n)$, let i_{n+1} be the smallest integer greater that i_n such that $|y_1(i_{n+1})| \ge |y_1(i_n)|$. Using the $\xi_1(i)$s and $y_1(i_j)$s we are now ready to construct Δ_1 through specifying its matrix kernel representation as follows:

$$\Delta_1 = \begin{pmatrix} \ddots & & & & & & & & & & \\ & \frac{\xi_1(i_1)}{y_1(i_1)} & & & & & & & & & \\ & \vdots & 0 & & & & & & & & \\ & \vdots & \vdots & \ddots & & & & & & & \\ & \frac{\xi_1(i_2-1)}{y_1(i_1)} & 0 & \cdots & 0 & & & & & & \\ & & & & & \frac{\xi_1(i_2)}{y_1(i_2)} & & & & & \\ & & & & & \vdots & 0 & & & & \\ & & & & & \vdots & \vdots & \ddots & & & \\ & 0 & & & & \frac{\xi_1(i_3-1)}{y_1(i_2)} & 0 & \cdots & 0 & & \\ & & & & & & & & & \frac{\xi_1(i_3)}{y_1(i_3)} & \\ & & & & & & & & & \vdots & \ddots \end{pmatrix}.$$

Notice that each row of the above matrix has at most one nonzero element, which, by the choice of the $y_1(i_j)$s, will have its absolute value less than or equal to one. This implies $\|\Delta_1\| \le 1$. Moreover, Δ_1 is clearly causal and it can easily be checked that $\Delta_1 y_1 = \xi_1$, which is what we wanted to show.

We now construct a nonlinear, time-invariant, and causal perturbation Δ_2. As before, Δ_2 must be such that $\|\Delta_2\| \le 1$ and $\Delta_2 y_2 = \xi_2$. Let Δ_2 be defined as follows:

$$(\Delta_2 f)(k) = \begin{cases} \xi_2(k-i) & \text{if for some integer} \\ & i \ge 0,\ P_k f = P_k S_i y_2 \\ 0 & \text{otherwise.} \end{cases}$$

Note that Δ_2 maps y_2 to ξ_2 and $\|\Delta_2\| \le 1$.

7.3. *Comparisons*

It is worthwhile comparing the class of perturbations that have gain less than unity over ℓ_2 (which arise in the standard μ) with the class of perturbations that have gain less than unity over ℓ_∞. If the perturbations are restricted to time-invariant ones, the ℓ_∞-stable perturbations with gain less than unity lie inside the unit ball of ℓ_2-stable perturbations (for the multivariable case, the unit ball will be scaled by a constant). This follows directly from the norm inequality between ℓ_1 and H_∞. If the perturbations are allowed to be time-varying, then the two sets are not comparable. Earlier, an example was presented that shows that the H_∞ ball is larger than the ℓ_1 ball. On the other hand, the operator Δ defined by

$$(\Delta f)(k) = f(0),$$

is ℓ_∞ stable but not ℓ_2 stable.

A question which might arise is, how do the derived robustness conditions differ from the Structured Singular Value? The answer lies in the class of perturbations assumed. While the perturbations here may be nonlinear time-varying (NLTV), nonlinear time-invariant (NLTI), or LTV for the conditions to be necessary and sufficient, μ theory gives necessary and sufficient conditions only for LTI perturbations. While the spectral radius test gives sufficient conditions for robustness for LTI perturbations it can be shown that in general these are not necessary, and thus it is potentially conservative when it is known that perturbations are LTI. In terms of computation, the robustness test proposed here is easy to compute and gives exact answers for any number of perturbation blocks, n. On the other hand, μ is much harder to compute especially since for $n > 3$ only an upper bound can be computed. One can use the Small Gain Theorem to get sufficient conditions for robust stability in the presence of NLTV ℓ^2 induced norm-bounded perturbations in the same way it was done for the **A** norm. In this case, a sufficient condition would be $\inf_{R \in \mathfrak{R}} \|R^{-1}\hat{M}R\|_{H_\infty} < 1$. It has recently been shown that it is also necessary (Shamma, 1992). This condition, however, is not sufficient to guarantee robustness when perturbations of the type considered in this paper are present, i.e. for ℓ_∞ induced norm-bounded perturbations. In contrast, robustness in the presence of ℓ_∞

TABLE 1. COMPARISONS BETWEEN DIFFERENT ROBUSTNESS CRITERIA

Perturbation class	$\mu(M) < 1$	$\inf_{R \in \mathcal{R}} \|R^{-1}MR\|_{H^\infty} < 1$	$\rho(\hat{M}) < 1$
NLTV, bounded ℓ_2-gain	nec	nec and suff	suff
NLTV, bounded ℓ_∞-gain	nec	nec	nec and suff
NLTI, bounded ℓ_2-gain	nec	suff	suff
NLTI, bounded ℓ_∞-gain	nec	nec	nec and suff
LTV, bounded ℓ_2-induced norm	nec	nec and suff	suff
LTV, bounded ℓ_∞-induced norm	nec	nec	nec and suff
LTI, bounded ℓ_2-induced norm	nec and suff	suff	suff
LTI, bounded ℓ_∞ induced norm	nec and suff	suff	suff

induced-norm bounded perturbations does imply robustness to ℓ_2 induced-norm bounded perturbations. The relationship between the various robustness conditions is summarized in Table 1 (In the table: nec, suff, respectively mean necessary and sufficient.)

Most of the proofs concerning μ are found in the tutorial paper by Packard and Doyle (1992) and in Shamma (1992) for the time varying case. The proof that μ is both necessary and sufficient in the case of LTI perturbations with bounded ℓ_1 norm goes as follows: consider the ℓ_∞ stability of $(I - M\Delta)^{-1}$ for all $\|\Delta\|_1 \leqq 1$. Suppose that $\hat{M}(e^{i\theta_0}) = c$ with $|c| \geq 1$. We need to show that there exists a LTI perturbation Δ with $\|\Delta\|_1 \leq 1$ such that $\hat{\Delta}(e^{i\theta_0}) = 1/c$. Let H be any system in ℓ_1 satisfying $\hat{H}(e^{i\theta_0}) = 1/c$ and $\hat{H}(e^{-i\theta_0}) = 1/\bar{c}$. The assertion can be verified by obtaining a solution to the model matching problem:

$$\inf_{Q \in \ell_1} \|H - (\lambda - e^{i\theta_0})(\lambda - e^{-i\theta_0})Q\|_1.$$

It can be shown that the solution of this problem has a value equal to $1/|c|$, and thus the existence of Δ is verified. The extension from the small gain theorem to μ is straightforward. It is also interesting to note that μ is both necessary and sufficient for general ℓ_p problems (Bamieh and Dahleh, 1993).

In terms of robust controller synthesis, the controller must be chosen such that $\rho(\hat{M})$ is minimized. The dependence of M on the controller is reflected through the Youla parameter, Q, since M can be expressed as (Francis (1987); Vidyasagar (1985); Youla *et al.* (1976))

$$M = M(Q) = T_1 - T_2QT_3,$$

where the T_is depend only on G_0. Because $\rho(\hat{M}) = \inf_{R \in \mathcal{R}} \|R^{-1}MR\|_1$, the robustness synthesis problem becomes one of finding

$$\inf_{Q \text{ stable}} \inf_{R \in \mathcal{R}} \|R^{-1}M(Q)R\|_1.$$

With Q stable and fixed, we have seen that picking the eigenvector associated with $\rho(\widehat{M(Q)})$ will yield the minimum value over all scalings in $\mathcal{R}$. When R is fixed, we have an ℓ_1-norm minimization problem. This problem and its solution will be discussed in the remaining part of the paper. So the approach which will be taken to solving the robustness synthesis problem is to start with an initial $R \in \mathcal{R}$. For that R we find the optimal Q resulting from the norm minimization problem. We then fix that Q and solve for the optimal R associated with this new Q and so on. Since at each step the objective function gets smaller and smaller, and since it is bounded from below by zero it is guaranteed to converge to some value. Unfortunately, this value may not be the global minimum. If at that point, a satisfactory level of performance robustness has been reached, we can stop and use the final Q to construct the controller. Otherwise, the iteration process should be restarted with a different initial scaling matrix in $\mathcal{R}$. This scheme is similar to the so-called D–K iteration used in the μ-synthesis technique (Doyle and Stein, 1981; Doyle, 1982). The main difference is that while the scales used for μ-synthesis are frequency dependent and a convex optimization problem must be solved at each frequency, the scalings here are not frequency dependent and can be readily found by computing the eigenvector associated with $\rho(\hat{M})$. Such a computation can be done very effectively using power methods, and no optimization problem need be solved to find the optimal scalings.

8 SYNTHESIS OF THE ℓ_1 CONTROLLER

As stated earlier, the ℓ_1 minimization problem is given by:

$$\mu_0 = \inf_{Q \text{ stable}} \|T_1 - T_2QT_3\|_1. \qquad \text{(OPT)}$$

In this section, we will show that this problem is equivalent to a linear programming problem in

infinite-dimensional space. By utilizing the duality theory of Lagrange multipliers, it is shown that in some cases the linear programs are in fact finite-dimensional and thus exact solutions for (OPT) can be obtained. For the rest of the cases, the duality theory provides upper and lower approximations of the optimal solution. The use of the Lagrange multiplier theory highlights the strong resemblance between the ℓ_1 problem and standard linear programming problems.

The admissible subspace $\mathcal{S}$ is defined as:

$$\mathcal{S} = \{R \in \ell_1^{m \times n} \mid R = T_1 Q T_3,\ Q \text{ is stable}\}.$$

The ℓ_1 problem can be interpreted as a distance problem: Find an element in the subspace $\mathcal{S}$ which is closest to the fixed element T_1, where distance is measured in the ℓ_1-norm. Previous work (Dahleh and Pearson, 1987, 1988a) used the duality theory for distance problems to arrive at a solution for (OPT). Here we take an alternate approach using Lagrange multiplier theory, which is in fact more intuitive and transparent, to arrive at similar conclusions.

8.1. *Characterization of the subspace* S

In the discussion below, it is assumed that $\hat{T}_2$ has full column rank $= n_2$, and $\hat{T}_3$ has full row rank $= n_3$. It is evident that this captures the most general situation since if either of these conditions does not hold, we can perform inner–outer factorizations on $\hat{T}_2$ and $\hat{T}_3$ and absorb the extra degree of freedom in $\hat{Q}$. Also, it is assumed that there exist n_2 rows of $\hat{T}_2$ and n_3 columns of $\hat{T}_3$ which are linearly independent for all λ on the unit circle. This assumption simplifies the exposition although it is not necessary. In general, it is enough to assume the above for one point on the unit circle (Staffans, 1990). Under this assumption, $\hat{T}_2$ and $\hat{T}_3$ can be written in the following form without loss of generality (possibly requiring the interchange of inputs and/or outputs):

$$\hat{T}_2 = \begin{pmatrix} \hat{T}_{21} \\ \hat{T}_{22} \end{pmatrix},$$

$$\hat{T}_3 = (\hat{T}_{31} \quad \hat{T}_{32}),$$

where $\hat{T}_{21}$ has dimensions $n_2 \times n_2$ and is invertible and $\hat{T}_{31}$ has dimensions $n_3 \times n_3$ and is invertible. Moreover, $\hat{T}_{21}$ and $\hat{T}_{31}$ have no transmission zeros on the unit circle. Thus $\hat{R} = \hat{T}_1 \hat{Q} \hat{T}_3$ can be written:

$$\hat{R} = \begin{pmatrix} \hat{T}_{21} \\ \hat{T}_{22} \end{pmatrix} \hat{Q} (\hat{T}_{31} \quad \hat{T}_{32}) = \begin{pmatrix} \hat{R}_{11} & \hat{R}_{12} \\ \hat{R}_{21} & \hat{R}_{22} \end{pmatrix}.$$

The objective is to obtain a characterization of the feasible set $\mathcal{S}$. Notice that $\hat{Q}$ can be uniquely determined from the equality $\hat{R}_{11} = \hat{T}_{21} \hat{Q} \hat{T}_{31}$. As was shown in Dahleh and Pearson (1987) and McDonald and Pearson (1991) the choice of $\hat{R}_{11}$ is constrained by the zeros of $\hat{T}_{21}$, $\hat{T}_{31}$ that are inside the unit disc. There is only a finite number of such zeros, and each zero is interpreted as a bounded linear functional on R_{11}. In the sequel, we use the following terminology.

Definition 1. A transfer function $\hat{G}$ interpolates $\hat{T}_{21}$, $\hat{T}_{31}$ if $\hat{T}_{21}^{-1} \hat{G} \hat{T}_{31}^{-1}$ is stable.

The motivation for this terminology stems from the fact that for $\hat{T}_{21}^{-1} \hat{G} \hat{T}_{31}^{-1}$ to be stable, $\hat{G}$ must have zeros at the same locations and directions as the zeros of $\hat{T}_{21}$ and $\hat{T}_{31}$. Each zero is in fact a bounded linear functional that annihilates the element G, and thus has a representation inside the dual space of ℓ_1, with the appropriate dimension. If these functionals are inside c_0, then we can view G as the annihilator in the dual of c_0. For example, let $\hat{G}(a) = 0$, where G is SISO, and $|a| < 1$. By definition of $\hat{G}$, we have $\hat{G}(a) = \sum_{k=0} g(k) a^k = 0$. Define $z_a = (1, a^2, a^3, \ldots) \in c_0$, then the interpolation condition can be expressed as $\langle z_a, G \rangle = 0$. If a is a complex number, then two functionals are defined, the real of z_a and the imaginary of z_a. The multivariable case carries more details, but the basic idea is the same (see Dahleh and Pearson, 1987; McDonald and Pearson, 1991).

The choice of $\hat{R}_{11}$ is constrained further so that the rest of the equations are still consistent, which in turn dictates a set of constraints on the rest of the elements of R. Define the following coprime polynomial factorizations:

$$\begin{aligned} \hat{T}_{22} \hat{T}_{21}^{-1} &= \hat{\tilde{D}}_2^{-1} \hat{\tilde{N}}_2, \\ \hat{T}_{31}^{-1} \hat{T}_{32} &= \hat{N}_3 \hat{D}_3^{-1}. \end{aligned} \tag{3.3}$$

Using these definitions, we state the following result characterizing the feasible set $\mathcal{S}$ for this case (McDonald and Pearson, 1991).

Theorem 3. Given $\hat{T}_2$, $\hat{T}_3$ with the assumptions as above, and $\hat{R} \in \mathbf{A}$, there exists $\hat{Q} \in \mathbf{A}$ satisfying $\hat{R} = \hat{T}_2 \hat{Q} \hat{T}_3$ if and only if:

(i) $$(-\hat{\tilde{N}}_2 \quad \hat{\tilde{D}}_2) \begin{pmatrix} \hat{R}_{11} & \hat{R}_{12} \\ \hat{R}_{21} & \hat{R}_{22} \end{pmatrix} = 0,$$

(ii) $$(\hat{R}_{11} \quad \hat{R}_{12}) \begin{pmatrix} -\hat{N}_3 \\ \hat{D}_3 \end{pmatrix} = 0,$$

(iii) $\hat{R}_{11}$ interpolates $\hat{T}_{21}$ and $\hat{T}_{31}$.

The conditions shown in parts (i), (ii) are convolution constraints on the ℓ_1 sequence R. The interpolation condition in the last part can be tightened, since only the common zeros of $\hat{T}_{21}$ and $\hat{T}_{22}$ need to be interpolated.

The discussion above shows that the characterization of $\mathcal{S}$ can be summarized by defining two operators,

$$\mathcal{V}: \ell_1^{m\times n} \to \mathbb{R}^s,$$

and

$$\mathcal{C}: \ell_1^{m\times n} \to \ell_1^r,$$

where s, r are some integers. The first operator captures the interpolation constraints, and thus has a finite dimensional range, and the second captures the convolution constraints. These two operators can be constructed in a straightforward fashion, bookkeeping being the only difficulty. To overcome this problem, it is helpful to think of R as a vector rather than a matrix. To illustrate this, let the operator $\mathcal{W}$ be a map from $\ell_1^{m\times n}$ to ℓ_1^{mn} defined as follows:

$$(\mathcal{W}R)(k): \begin{pmatrix} r_{11}(k) \\ \\ r_{m1}(k) \\ r_{21}(k) \\ \\ r_{mn}(k) \end{pmatrix}$$

The operator $\mathcal{W}$ is a one-to-one and onto operator, whose inverse is equal to its adjoint (a fact used later). It simply rearranges the variables in R. The conditions on R presented in the above theorem can be written explicitly in terms of each component of R.

To construct the first operator $\mathcal{V}$, recall that each interpolation condition is interpreted as a bounded linear functional on R. By stacking up these functionals, the operator $\mathcal{V}$ is constructed. For example, suppose $\hat{T}_{21}$ and $\hat{T}_{31}$ are SISO and both have N zeros a_i in the open unit disc. Then the matrix $\mathcal{V}$ is given by $\mathcal{V} = V_\infty \mathcal{W}$ where

$$V_\infty = \begin{pmatrix} \mathrm{Re}\,(a_1^0) & 0 & 0 & 0 & \mathrm{Re}\,(a_1^1) & 0 & \cdots & \mathrm{Re}\,(a_1^j) & 0 & \cdots \\ \mathrm{Im}\,(a_1^0) & 0 & 0 & 0 & \mathrm{Im}\,(a_1^1) & 0 & \cdots & \mathrm{Im}\,(a_1^j) & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \vdots & \vdots & \vdots \\ \mathrm{Re}\,(a_N^0) & 0 & 0 & 0 & \mathrm{Re}\,(a_N^1) & 0 & \cdots & \mathrm{Re}\,(a_N^j) & 0 & \cdots \\ \mathrm{Im}\,(a_N^0) & 0 & 0 & 0 & \mathrm{Im}\,(a_N^1) & 0 & \cdots & \mathrm{Im}\,(a_N^j) & 0 & \cdots \end{pmatrix} \quad j = 0, 1, 2, \ldots$$

For the second operator, $\mathcal{C}$, recall that convolution can be interpreted as multiplication by a block Toeplitz matrix, in this case with finite memory since $\tilde{N}_2$, $\tilde{D}_2$, N_3 and D_3 all have finite length (the corresponding λ-transform is a polynomial). By simple rearrangement, the operator is constructed with its image inside ℓ_1^r. Hence $\mathcal{C}$ is given by $\mathcal{C} = \mathcal{T}\mathcal{W}$, where $\mathcal{T}$ is a block lower triangular matrix. For a detailed example (see Dahleh and Pearson, 1988a; McDonald and Pearson, 1991).

To illustrate the construction of the operator $\mathcal{T}$, consider as an example the coprime-factor perturbation problem considered earlier for a SISO. The condition for stability robustness is given by Dahleh (1992)

$$\|[\tilde{V} - Q\tilde{N} \quad -\tilde{U} + Q\tilde{M}]\|_1 \leq 1.$$

where $\tilde{V}$, $\tilde{U}$ are left coprime factors of a nominal controller. In this case, $T_2 = 1$ and $T_3 = (\tilde{N} \quad -\tilde{M})$. Since $\tilde{M}^{-1}\tilde{N} = NM^{-1}$ with N, M coprime, the conditions in the above theorem translate to

$$(R_{11} \quad R_{12})\begin{pmatrix} M \\ N \end{pmatrix} = 0.$$

The matrix $\mathcal{T}$ is then given by:

$$\mathcal{T} = \begin{pmatrix} (m(0) & n(0)) & 0 & & 0 & & 0 & & 0 & \cdots \\ (m(1) & n(1)) & & n(0)) & 0 & & 0 & & 0 & \cdots \\ (m(2) & n(2)) & & n(1)) & (m(0) & n(0)) & 0 & & 0 & \cdots \\ (m(3) & n(3)) & & n(2)) & (m(1) & n(1)) & (m(0) & n(0)) & 0 & \cdots \\ & \vdots & & & & & \vdots & & \vdots & \vdots \end{pmatrix}$$

It is interesting to note that in this example the operator $\mathcal{C}$ captures all the conditions and no interpolation conditions are needed. The conditions presented in the theorem can be redundant, and can be significantly reduced (Staffans, 1990).

The subspace $\mathcal{S}$ is then the set of all elements $R \in \ell_1^{m\times n}$ such that $\mathcal{V}R = 0$ and $\mathcal{C}R = 0$. Let

$b_1 = \mathcal{V}T_1$, $b_2 = \mathcal{C}T_1$, and $\Phi = T_1 - R$. The ℓ_1 optimization problem can be restated as:

$$\inf_{\Phi \in \ell_1^{m\times n}} \|\Phi\|_1 \quad \text{subject to} \quad \mathcal{V}\Phi = b_1, \quad \mathcal{C}\Phi = b_2, \tag{OPT}$$

8.2. *Relations to linear programming*

It is well-known that in finite-dimensional spaces ℓ_1-norm minimization is equivalent to linear programming. This turns out to be true in general, and can be justified as follows: Let $\Phi = \Phi^1 - \Phi^2$, with $\phi_{ij}^1(k)$, $\phi_{ij}^2(k) \geq 0$. The norm is then replaced by the function $\max_i \sum_{j,k} \phi_{ij}^1(k) + \phi_{ij}^2(k)$. Define the operator $\mathcal{N}: \ell_1^{m\times n} \to \mathbb{R}^m$ by $(\mathcal{N}\Phi)_i = \sum_{j,k} \phi_{ij}(k)$. The following problem is easily seen to be equivalent to (OPT):

$$\inf \mu,$$

subject to

$$\begin{aligned} \mathcal{N}(\Phi^1 + \Phi^2) - \mu e &\leq 0, \\ \mathcal{V}(\Phi^1 - \Phi^2) &= b_1, \\ \mathcal{C}(\Phi^1 - \Phi^2) &= b_2, \\ \phi_{ij}^1(k), \ \phi_{ij}^2(k) &\geq 0, \end{aligned}$$

where $e \in \mathbb{R}^m$ and $e^T = (1, 1, \ldots, 1)$. It is interesting to notice that if Φ^1, Φ^2 where restricted to finite impulse response sequences, the above problem is readily a linear programming problem. This will turn out to be a crucial observation in obtaining approximate solutions, as will be described later on.

8.3. *Lagrange multiplier formulation*

Let $X = \ell_1^{m\times n} \times \ell_1^{m\times n} \times \mathbb{R}$ and $Z = \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s \times \ell_1^r \times \ell_1^r$. Let P_X, P_Z denote the positive cones inside X, Z consisting of elements with nonnegative pointwise components. Define the operator $\mathcal{A}: X \to Z$, decomposed conformally with X and Z, and the vector $b \in Z$ as follows:

$$\begin{pmatrix} \mathcal{N} & \mathcal{N} & -e \\ \mathcal{V} & -\mathcal{V} & 0 \\ -\mathcal{V} & \mathcal{V} & 0 \\ \mathcal{C} & -\mathcal{C} & 0 \\ -\mathcal{C} & \mathcal{C} & 0 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ b_1 \\ -b_1 \\ b_2 \\ -b_2 \end{pmatrix}$$

Define the linear functional $c^* = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ on X.

With these definitions, (OPT) becomes:

$$\inf \langle x, c^* \rangle,$$

subject to

$$Ax \leq b,$$
$$x \in X, \quad x \geq 0,$$

where $x \in X$ has the form

$$\begin{pmatrix} \Phi^1 \\ \Phi^2 \\ \mu \end{pmatrix}$$

All the inequalities should be interpreted with respect to the positive cones. It is interesting that with the above definitions, (OPT) looks very much like a standard linear programming problem, with the exception that the number of variables and constraints is infinite.

The Lagrange multiplier is an element inside Z^*, the dual space of Z which can be identified as: $Z^* = \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s \times c_0^r \times c_0^r$. (Here we have assumed that Z is equipped with the weak* topology, not the norm topology.) The dual cone P_Z^* again consists of the nonnegative elements in Z^*. The Lagrangian can be defined as

$$\begin{aligned} L(x, z^*) &= \{\langle x, c^* \rangle + \langle \mathcal{A}x - b, z^* \rangle\} \\ &= \{\langle x, c^* + \mathcal{A}^* z^* \rangle - \langle b, z^* \rangle\}, \end{aligned}$$

where $\mathcal{A}^*: Z^* \to X^*$ is the adjoint operator of $\mathcal{A}$. From the theory of Lagrange multipliers (Luenberger, 1969), the minimum solution can be obtained by performing an unconstrained minimization of L, i.e.

$$\mu_0 = \sup_{z^* \geq 0} \inf_{x \geq 0} \{\langle x, c^* + \mathcal{A}^* z^* \rangle - \langle b, z^* \rangle\}.$$

This result is true despite the fact that the constraints do not satisfy the regularity conditions (Dahleh *et al.* (1993)). Clearly for μ_0 to be finite, i.e. $\mu_0 > -\infty$, $c^* + \mathcal{A}z^* \geq 0$ and hence the above infimization is achieved for $x = 0$. This gives a dual formulation of (OPT) summarized as:

$$\mu_0 = \sup_{z^* \geq 0} \langle b, -z^* \rangle$$
$$\text{subject to} \quad c^* + \mathcal{A}^* z^* \geq 0. \tag{DOPT}$$

To evaluate this explicitly, let $\mathcal{A}^*$, z^* be given by:

$$\mathcal{A}^* = \begin{pmatrix} \mathcal{N}^* & \mathcal{V}^* & -\mathcal{V}^* & \mathcal{C}^* & -\mathcal{C}^* \\ \mathcal{N}^* & -\mathcal{V}^* & \mathcal{V}^* & -\mathcal{C}^* & \mathcal{C}^* \\ -e^T & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$z^* = \begin{pmatrix} \eta \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

By direct substitution, (DOPT) is converted to

$$\mu_0 = \sup \langle b_1, \alpha_1 - \alpha_2 \rangle + \langle b_2, \beta_1 - \beta_2 \rangle,$$

subject to

$$\mathcal{N}^*\eta + \mathcal{V}^*(\alpha_1 - \alpha_2) + \mathcal{C}^*(\beta_1 - \beta_2) \geq 0,$$
$$\mathcal{N}^*\eta - \mathcal{V}^*(\alpha_1 - \alpha_2) - \mathcal{C}^*(\beta_1 - \beta_2) \geq 0,$$
$$\sum_{i=1}^{m} \eta_i \leq 1,$$
$$\alpha_1, \alpha_2, \beta_1, \beta_2, \eta \geq 0.$$

Finally, substituting $\alpha = \alpha_1 - \alpha_2$ and $\beta = \beta_1 - \beta_2$ we get

$$\mu_0 = \sup \langle b_1, \alpha\rangle + \langle b_2, \beta\rangle,$$

subject to

$$-\mathcal{N}^*\eta \leq \mathcal{V}^*\alpha + \mathcal{C}^*\beta \leq \mathcal{N}^*\eta,$$
$$\sum_{i=1}^{m} \eta_i \leq 1, \quad \eta \geq 0, \qquad \text{(DOPT)}$$
$$\alpha \in \mathbb{R}^{v}, \quad \beta \in c_0^r.$$

This dual formulation sheds a new light on the optimization problem. In our context, it will provide two important results: the existence of finite-dimensional duals for specific classes of problems, and the ability to construct suboptimal solutions that are within a prescribed ϵ from the actual minimum.

Comment.

The computation of the adjoint operators is quite simple once the operators are already constructed. Recall that $\mathcal{V} = V_\infty \mathcal{W}$; hence the adjoint operator $\mathcal{V}^* = \mathcal{W}^{-1}V_\infty^T$. Similarly, $\mathcal{C}^* = \mathcal{W}^{-1}\mathcal{T}^T$. Matrix representations of the operator $\mathcal{N}$ and its adjoint are obtained in a similar fashion.

8.4. *Exact solutions for a class of problems*

Let the space $\mathcal{S}$ be characterized solely by interpolation conditions. This is the situation when both $\hat{T}_2$ and $\hat{T}_3$ have full row rank and column rank, respectively. In this case $\mathcal{C} = 0$ and $b_2 = 0$. The dual problem (DOPT) involves only a finite number of variables and thus it is a finite-dimensional problem. The constraints however are infinite. Since the elements of $\mathcal{V}^*$ were constructed from zeros inside the unit disc, the entries will eventually decay and only a finite number of the constraints are active. A bound on the number of such constraints can be derived (Dahleh and Pearson, 1987). The problem is now a standard finite-dimensional linear program, which can be solved exactly. The solution to the primal problem (OPT) can be constructed either by the alignment conditions, or by observing that the dual of (DOPT) is exactly the primal problem.

8.5. *Approximation*

In the sequel, we will assume that $\mathcal{C}T_1 = b_2$ is a finite impulse response sequence. This condition is equivalent to saying that there exists a FIR feasible solution for (OPT). If this condition is not satisfied, then the problem can be modified so that the condition will hold (Dahleh and Pearson, 1988a; McDonald and Pearson, 1991). Upper approximations of μ_0 can be readily obtained from the primal problem. Define $\bar{\mu}_N$ as follows:

$$\bar{\mu}_N = \min \mu,$$

subject to

$$\mathcal{N}(\Phi^1 + \Phi^2) - \mu e \leq 0,$$
$$\mathcal{V}(\Phi^1 - \Phi^2) = b_1,$$
$$\mathcal{C}(\Phi^1 - \Phi^2) = b_2,$$
$$\phi^1_{ij}(k), \quad \phi^2_{ij}(k) \geq 0,$$
$$\phi^1_{ij}(k) = 0, \quad \phi^2_{ij}(k) = 0 \,\forall k > N.$$

Since $\mathcal{C}$ is constructed from FIR sequences, this optimization will involve a finite number of variables and a finite number of constraints. It is evident that $\bar{\mu}_N$ is a non-increasing sequence satisfying $\mu_0 \leq \bar{\mu}_N$ for all N. Also, since feasible FIR solution exists, then $\bar{\mu}_N$ is finite for N large enough. Since FIR solutions are dense, it follows that $\bar{\mu}_N \to \mu_0$ as $N \to \infty$. For each $\bar{\mu}_N$ a solution for the primal problem can be constructed. The difficulty with this procedure is that it is not clear how far the solution is from optimal at any given N. This will be overcome by presenting lower approximations of the problem.

It is interesting to notice that the dual of this problem is obtained through truncating the constraints of the dual problem (DOPT). Another approximation obtained from the dual problem can be obtained by truncating the variables $\beta \in c_0^r$ (Dahleh, 1992; Staffans, 1990). Define $\underline{\mu}_N$ as follows:

$$\underline{\mu}_N = \max \langle b_1, \alpha\rangle + \langle b_2, \beta\rangle,$$

subject to

$$-\mathcal{N}^*\eta \leq \mathcal{V}^*\alpha + \mathcal{C}^*\beta \leq \mathcal{N}^*\eta,$$
$$\sum_{i=1}^{m} \eta_i \leq 1, \; \eta \geq 0,$$
$$\alpha \in \mathbb{R}^{v}, \quad \beta \in c_0^r, \quad \beta(k) = 0 \,\forall k > N.$$

It is evident that $\underline{\mu}_N \leq \mu_0$ and that $\underline{\mu}_N \to \mu_0$ as $N \to \infty$. The former assertion is due to the fact that the new problem has fewer degrees of freedom, and the latter is due to the fact that finite sequences are dense in c_0. The above problem is not immediately a finite-dimensional problem—the constraints due to the operator $\mathcal{V}^*$ are still infinite; however, only a finite subset of

these are active as it was in the case where $\mathscr{C}$ was equal to 0. A complete discussion of the computation of this problem is given in Staffans (1990). Clearly, there is no feasible solution for the primal problem for any of the μ_Ns.

8.6. *Computations*

In the case where $\mathscr{C} = 0$, the ℓ_1 minimization problem is solved exactly. In all other cases, only approximate solutions are obtained through obtaining upper and lower approximations of μ_0. A third method of solution known as the Delay Augmentation method was proposed in Diaz-Bobillo and Dahleh (1992). The basic idea in there is to augment the matrices T_2 and T_3 with delays so that they become square matrices. A square problem is then solved as the number of delays is increased. This method has the following definite advantages over the methods proposed earlier: (1) it involves solving one linear program at each iteration that provides upper and lower bounds and feasible solutions, (2) it only requires computing interpolation conditions, which can be done using matrix computation, (3) it captures in a precise way the exact order of the optimal controller, and does not cause order inflation due to approximations, (4) in many special cases, the exact optimal controller can be computed for the non-square problem.

The general Lagrange Multiplier approach allows the above theory to be extended to handle arbitrary problems with linear constraints, via solving linear programming problems. Details on exact and approximate solutions of such problems including mixed H_∞ and ℓ_1 problems are reported in Dahleh *et al.* (1992).

To obtain fast solutions that do not necessarily capture the structure of this problem, one can follow the approach in Boyd and Barratt (1991) in which one seeks direct FIR solutions for Q. This problem can be posed as a linear programming problem which can approximate the actual solution arbitrarily closely. However, unless one invokes duality, the difference between the approximate and actual value of μ_0 remains unknown. No information can be inferred about the structure of the optimal controller.

It is interesting to note that exact solutions for special problems with $\mathscr{C} \neq 0$ have been constructed in Diaz-Bobillo and Dahleh (1993, 1992) and Staffans (1991). Although existence of ℓ_1-optimal solutions is guaranteed (under mild conditions, namely no interpolations on the unit circle), it is not known whether these solutions are rational or not. If $\mathscr{C} = 0$ optimal solutions are FIR, and hence rational. The general case is still an active area of research.

9 CONCLUSIONS

This paper gives an overview of the problem of synthesizing optimal controllers to deliver performance specifications in the time domain, in the presence of plant uncertainty and bounded but unknown exogenous inputs. A general framework for the robust performance problem is presented from which necessary and sufficient conditions are derived. These conditions were related to the spectral radius of a matrix constructed from the configuration of the closed-loop system. Alternate equivalent conditions are also discussed in terms of linear matrix inequalities. These conditions are in turn used in the synthesis problem, which requires the solution of an ℓ_1 optimal control problem. A solution of this problem using the duality theory of Lagrange multipliers is used. This approach highlights in a non-trivial way the relations between ℓ_1 optimization problems for infinite-dimensional systems and infinite linear programming problems. In fact, the solutions presented exploit the problem structure and do not rely on a general theory for solving infinite linear programming problems, since such a theory does not exist.

This paper discusses only discrete-time problems. The interest in discrete-time systems stems from the fact that most controllers these days are digital controllers and are interfaced with the continuous-time plant through A/D and D/A converters. A better formulation should have a hybrid system consisting of both continuous- and discrete-time dynamics. Such systems have recently received considerable attention from the control community and are known as sampled-data systems. A formulation of the ℓ_1 sampled-data problem can be found in Bamieh *et al.* (1992), Dullerud and Francis (1992), Khammash (1992) and Sivashankar and Khargonekar (1991) in which it is shown that synthesizing a digital controller for a continuous-time plant can be done by solving a purely discrete-time problem. This motivates the earlier discussion.

There are other related problems that are not discussed in this paper. The problem of designing controllers for tracking a specific trajectory is an important problem and was solved in Dahleh and Pearson (1988b). The problem of finding conditions for the robustness of time-varying system in the presence of structured uncertainty has been solved in Khammash (1992). The problem of determining the merits of time-varying compensation vs

time-invariant compensation was studied in Khammash and Dahleh (1993). The ℓ_1 synthesis approach has been extended for periodic and multi-rate sampled plants (Dahleh *et al.*, 1992). Also, this theory was successfully incorporated as part of an adaptive control scheme, in which the stability of the closed loop system was guaranteed for a larger set of plant uncertainty (Dahleh and Dahleh, 1990; Voulgaris *et al.*, 1993). Finally, a case-study for the applicability of this theory was reported in Dahleh and Richards (1989) in which a ℓ_1 controller was designed for a model of the $X-29$ aircraft.

A pressing research problem is the understanding of the structure of the optimal ℓ_1 controllers. Such an understanding will not only add insight into the problem, but will also offer simpler ways of computing the optimal solution. This has been the case for the H_∞ and H_2 problems. Some interesting results in that direction are reported in Staffans (1991) in which exact solutions for the infinite-dimensional linear programs arising in some special non-square problems have been computed. Also, it was shown in Diaz-Bobillo and Dahleh (1991) that optimal solutions may require a dynamic controller even though all the states are available. The existence of some separation structure on the ℓ_1 problem (similar to that of the H_∞ problem (Doyle *et al.*, 1989)) is still under investigation.

Another important research direction is the synthesis problem by exactly minimizing the spectral radius function, rather than the iterative scheme suggested. The iterative scheme is guaranteed to converge only to local minima and hence there is a need for looking for another approach for minimizing this function.

In this paper, a comparison between the spectral radius function and μ is sketched. At this point, it is not known whether there exist examples in which the two methods exhibit extreme behavior. Research in that direction is currently in progress.

REFERENCES

Bamieh, B. A. and M. Dahleh (1993). On robust stability with structured time-invariant perturbations. *Syst. Contr. Lett.* (Submitted.)

Bamieh, B. A., M. A. Dahleh and J. B. Pearson (1992). Minimization of the L^∞-induced norm for sampled-data systems. *IEEE Trans. Aut. Control.* (To appear.)

Boyd, S. P. and C. H. Barratt (1991). *Linear Controller Design: Limits of Performance.* Prentice Hall, Englewood Cliffs, NJ.

Boyd, S. P. and J. C. Doyle (1987). Comparison of peak and RMS gains for discrete-time systems. *Syst. Contr. Lett.*, **9**, 1–6.

Dahleh, M. A. (1992). BIBO stability robustness for coprime factor perturbations. *IEEE Trans. Aut. Control*, **37**, 352–355.

Dahleh, M. and M. A. Dahleh (1990). Optimal rejection of persistent and bounded disturbances: continuity properties and adaptation. *IEEE Trans. Aut. Control*, **35**, 687–696.

Dahleh, M. A. and J. B. Pearson (1987). ℓ^1 optimal feedback controllers for MIMO discrete-time systems. *IEEE Trans. Aut. Control*, **32**, 314–322.

Dahleh, M. A. and J. B. Pearson (1988a). Optimal rejection of persistent disturbances, robust stability and mixed sensitivity minimization. *IEEE Trans. Aut. Control*, **33**, 722–731.

Dahleh, M. A. and J. B. Pearson (1988b). Minimization of a regulated response to a fixed input. *IEEE Trans. Aut. Control*, **33**, 924–930.

Dahleh, M. A. and Y. Ohta (1988). A necessary and sufficient condition for robust BIBO stability *Syst. Contr. Lett.*, **11**, 271–275.

Dahleh, M. A. and D. Richards (1989). Application of modern control theory on a model of the X-29 aircraft, LIDS, MIT, report No. LIDS-P-1932.

Dahleh, M. A., P. Voulgaris and L. Valavani (1992). Optimal and robust controllers for periodic and multi-rate systems. *IEEE Trans. Aut. Control*, **37**, 90–99.

Dahleh, M. A., N. Elia and I. Diaz-Bobillo (1993). Controller design via linear programming. (In preparation.)

Deodhare, G. and M. Vidyasagar (1990). Some results on ℓ_1-optimality of feedback control systems: the SISO discrete-time case. *IEEE Trans. Aut. Control*, **35**, 1082–1085.

Desoer, C. A. and M. Vidyasagar (1975). *Feedback Systems: Input–Output Properties.* Academic, NY.

Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainty. *IEE Proceedings*, **129**, 242–250.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE-Trans. A-C*, **26**, 4–16.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Aut. Control*, **34**, 831–847.

Diaz-Bobillo, I. and M. A. Dahleh (1993). State feedback ℓ_1 optimal controllers can be dynamic, LIDS, MIT, Report No. LIDS-P-2051. *Systems and Control Letters.* (To appear.)

Diaz-Bobillo, I. and M. A. Dahleh (1992). Minimization of the maximum peak-to-peak gain: the general multiblock problem. *IEEE Trans. Aut. Control.* (To appear).

Dullerud, G. and B. A. Francis (1992). $\mathcal{L}^1$ analysis and design of sampled-data systems. *IEEE Trans. Aut. Control*, **37**, 436–446.

Francis, B. A. (1987). *A Course in* H_∞ *Control Theory.* Springer-Verlag, Berlin.

Glover, K. and D. McFarlane (1989). Robust stabilization of normalized coprime factor plant description with H_∞ bounded uncertainty. *IEEE Trans. Aut. Control*, **34**, 831–830.

Horn, R. and C. Johnson (1985). *Matrix Analysis.* Cambridge University Press.

Khammash, M. (1992). Necessary and sufficient conditions for the robustness of time-varying systems with applications to sampled-data systems. *IEEE Trans. Aut. Control.* (To appear.)

Khammash, M. and J. B. Pearson (1990). Robust disturbance rejection in ℓ^1-optimal control systems. *Syst. Contr. Lett.*, **14**, 93–101.

Khammash, M. and M. Dahleh (1993). Time-varying control and the robust performance of systems with structured norm-bounded uncertainty. *Automatica.* (To appear.)

Khammash, M. and J. B. Pearson (1991). Performance robustness of discrete-time systems with structured uncertainty. *IEEE Trans. Aut. Control*, **36**, 398–412.

Khammash, M. and J. B. Pearson (1991). Robustness synthesis for discrete-time systems with structured uncertainty. *Proc. of* 1991 *ACC*, 2720–2724

Khammash, M. and J. B. Pearson (1992). Analysis and design of robust performance with structured uncertainty. *Syst. Contr. Lett.* (To appear.)

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. John Wiley, NY

McDonald, J. S. and J. B. Pearson (1991) ℓ_1-optimal control of multivariable systems with output norm constraints. *Automatica*, **27**, 317–329.

Mendlovitz, M. A. (1989). A simple solution to the ℓ_1 optimization problem. *Syst. Contr. Lett.*, **12**, 461–463.

Packard, A. and J. C. Doyle (1992) The complex structured singular value. *Automatica*. (To appear.)

Shamma, J (1992). Robust stability with time-varying structured uncertainty. *IEEE Trans. Aut. Control*. (Submitted.)

Sivashankar, N. and P. Khargonekar (1991). $\mathscr{L}_\infty$-induced norm of sampled-data systems. *Proc* 1991 *CDC*.

Staffans, O. J. (1990). On the four-block model matching problem in ℓ_1. Helsinki University of Technology, Espoo, Report A289.

Staffans, O. J. (1991). Mixed sensitivity minimization problems with rational ℓ_1-optimal solutions. *J. of Optimization Theory and Applications*, **70**, 173–189.

Vidyasagar, M. (1981). Input–output analysis of large-scale

Vidyasagar, M. (1981). Input–output analysis of large-scale interconnected systems. *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin.

Vidyasagar, M. (1985). *Control Systems Synthesis: A Factorization Approach*. MIT press, Cambridge, MA.

Vidyasagar, M. (1986). Optimal rejection of persistent bounded disturbances. *IEEE Trans. Aut. Control*, **31**, 527–534.

Voulgaris, P., M. A. Dahleh and L. Valavani (1993). Slowly-varying systems: ℓ_∞ to ℓ_∞ performance implications to adaptive control. *Automatica*. (Submitted.)

Willems, J. C. (1971). *The Analysis of Feedback Systems*. MIT Press, Cambridge, MA.

Youla, D. C., H. A. Jabr and J. J. Bongiorno (1976). Modern Wiener–Hopf design of optimal controllers—part 2: the multivariable case. *IEEE Trans. Aut. Control*, **21**, 319–338.

Zames, G. (1981). Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Trans. Aut. Control*, **26**, 301–320.

Automatica, Vol. 29, No. 1, pp. 57–70, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ Control for Discrete-time Systems via Convex Optimization*†

ISAAC KAMINER,‡ PRAMOD P. KHARGONEKAR‡ and MARIO A. ROTEA§

A mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem for discrete-time systems is solved by converting it into a convex optimization problem over a finite-dimensional space.

Key Words—Robust control; multiobjective control; convex programming.

Abstract—A mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control problem for discrete-time systems is considered, where an upper bound on the $\mathcal{H}_2$ norm of a closed loop transfer matrix is minimized subject to an $\mathcal{H}_\infty$ constraint on another closed loop transfer matrix. Both state-feedback and output-feedback cases are considered. It is shown that these problems are equivalent to finite-dimensional convex programming problems. In the state-feedback case, nearly optimal controllers can be chosen to be static gains. In the output feedback case, nearly optimal controllers can be chosen to have a structure similar to that of the central single objective $\mathcal{H}_\infty$ controller. In particular, the state dimension of nearly optimal output-feedback controllers need not exceed the plant dimension.

1. INTRODUCTION

DESIGN OF CONTROL systems almost invariably involves tradeoffs among competing objectives. It is often the case that the controller is required to meet several different performance and robustness goals, and all of these cannot be met simultaneously. For example, it is intuitively clear that to obtain a greater robust stability margin, it is likely that the performance of the control system needs to be compromised. In classical single loop feedback design, these tradeoffs are performed in terms of the (open) loop transfer function. For instance, stability margins in terms of either the phase/gain margins or the distance of the Nyquist plot to the critical point are traded off against disturbance rejection at low frequencies. Clearly, it is important to develop analytical tools to help the designer understand how the various competing objectives conflict with each other. From this point of view, one should postulate the controller synthesis problem as the problem of studying tradeoffs among competing objectives. For a more detailed discussion of multiobjective controller synthesis as well as additional references, see Boyd and Barratt (1990), Dorato (1991), Khargonekar and Rotea (1991b), and Rotea (1990).

The subject of this paper is a certain constrained optimal controller synthesis problem—the so-called mixed $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis problem. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems can be motivated in many different ways. As a matter of fact, there are many *different* mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems. These problems are one way of analytically formulating the issue of tradeoffs in control system synthesis.

To give a brief description of the various mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems, consider the feedback system shown in Fig. 1. Let $T_{z_iw_i}$, $i = 0, 1$, denote the closed loop transfer matrix from the exogenous input w_i to the controlled output z_i. One mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem is to find an internally stabilizing controller $\mathscr{C}$ which minimizes $\|T_{z_0w_0}\|_2$ subject to the constraint $\|T_{z_1w_1}\|_\infty < \gamma$. This problem is equivalent to a problem of optimal nominal performance subject to a robust stability requirement. To be more specific, a controller that solves this mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem will ensure that the closed loop system is robustly stable to all finite-gain stable (possibly nonlinear time-varying) perturbations Δ, interconnected to the system by $w_1 = \Delta z_1$, such that $\|\Delta\|_\infty \le 1/\gamma$. On the other hand, $\|T_{z_0w_0}\|_2$ represents the steady-state variance of the output z_0 when $w_1 = 0$ and w_0 is white noise with unit intensity. Currently no analytic solution to this mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem is available. Rotea and Khargonekar (1991a) have

* Received 30 August 1991; received in final form 7 April 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor B. A. Francis.

† Supported in part by NSF under grant no. ECS-9096109 and RIA no. ECS-91-08493, AFOSR under contract no. AFOSR-90-0053, ARO under grant no. DAAL03-90-G-0008.

‡ Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, U.S.A.

§ School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907-1282, U.S.A.

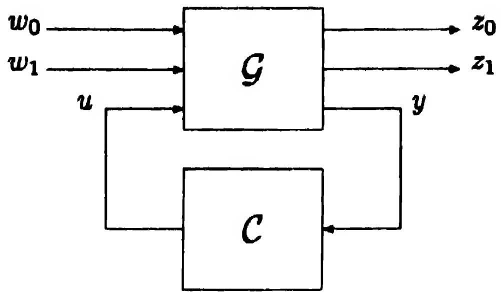


FIG. 1.

obtained some sufficient conditions for the solvability of this mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem in the state-feedback case. While the results in Rotea and Khargonekar (1991a) have been obtained for continuous-time systems, many of them can be extended quite easily to discrete-time systems.

A somewhat different mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem formulation was introduced by Bernstein and Haddad (1989) to combine the LQG and $\mathcal{H}_\infty$ controller design theories. This problem is restricted to the case $w_0 = w_1 =: w$. Instead of minimizing $\|T_{z_0w}\|_2$, they considered the minimization of an "upper bound" for $\|T_{z_0w}\|_2$, subject to the constraint $\|T_{z_1w}\|_\infty < \gamma$. Recently, many papers have appeared that address this mixed $\mathcal{H}_2/\mathcal{H}_\infty$ controller design problem, see for example, Bambang et al. (1990), Doyle et al. (1989a), Khargonekar and Rotea (1991a, b), Mustafa and Bernstein (1991), Steinbuch and Bosgra (1991), Yeh et al. (1992), Zhou et al. (1990) and the references cited therein.

The mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem considered in this paper not only provides a more tractable approach to the problem of minimizing nominal performance subject to a robust stability constraint, but it can also be interpreted as an optimal performance problem. Indeed, as shown by Zhou et al. (1990) in the continuous-time case, the "dual" of the auxiliary cost or upper bound of Bernstein and Haddad (1989) is closely related to a system gain from a combination of power and white noise exogenous inputs to the power of the regulated output.

In this paper, we focus on the discrete-time version of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem as formulated by Bernstein and Haddad (1989). While much work has been done on the continuous-time case, the discrete-time case has received much less attention. Indeed, at this time the discrete-time analog of the coupled Riccati equations obtained by Bernstein and Haddad (1989) for the continuous-time case are not available for output-feedback problems. Some results for discrete-time mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems have been obtained by Bambang et al. (1990), Haddad et al. (1991), and Mustafa and Bernstein (1991). Mustafa and Bernstein (1991) have considered the static state-feedback problem and derived sufficient conditions for optimality of a state-feedback gain. Bambang et al. (1990) and Haddad et al. (1991) have considered the static output-feedback problem. It seems that at this time no solution to the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem is available in the general dynamic output-feedback case. This is the primary motivation for this paper.

Our approach is as the recent paper by Khargonekar and Rotea (1991a) where a convex optimization approach to the continuous-time mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem introduced in Bernstein and Haddad (1989) has been developed. A similar approach has been applied earlier by Bernussou et al. (1989) to a quadratic stability problem. The starting point is to take a "sub-optimal approach". More specifically, with J denoting the "mixed $\mathcal{H}_2/\mathcal{H}_\infty$" performance measure (a precise definition of J is given in Section 2) let

$$\nu(\mathcal{G}) := \inf \{J : \mathcal{C} \text{ internally stabilizing and } \|T_{z_1w}\|_\infty < \gamma\},$$

denote the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure. Then we consider the following problem:

> "Compute $\nu(\mathcal{G})$ and given $\alpha > \nu(\mathcal{G})$, find an internally stabilizing controller $\mathcal{C}$ such that $\|T_{z_1w}\|_\infty < \gamma$, and the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure satisfies $J < \alpha$".

The main results of this paper are contained in Sections 4 and 5. The full-information/state-feedback case is considered in Section 4, while the output-feedback case is considered in Section 5. It is shown that if the plant state is available for feedback, one can come arbitrarily close to the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure using constant gain (i.e. nondynamic) state-feedback controllers. In other words, in the state-feedback case, static gain controllers offer the best possible performance.

In the full-information feedback case (i.e. both the exogenous input and the system state are available for feedback) there is a significant departure from the continuous-time case. It turns out that in the discrete-time case, one cannot come arbitrarily close to the infimum by taking static state-feedback controllers. The best that one can do is to use static full-information controllers. As a consequence, this result is of little practical interest except that it is critically useful in dealing with the output-feedback case. This situation is similar to that in the single objective standard $\mathcal{H}_\infty$ control problem for

discrete-time systems as in Basar and Bernhard (1991), Iglesias and Glover (1991), Limebeer *et al.* (1989), Liu *et al.* (1991) and Stoorvogel (1990).

It is shown that in the state-feedback as well as the full-information case, the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ optimal performance $\nu(\mathcal{G})$ and a static gain controller that satisfies $J < \alpha$ (for any $\alpha > \nu(\mathcal{G})$) can be obtained by solving a finite-dimensional convex programming problem over a bounded set of real matrices.

In the output-feedback case, it is shown that the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control problem can be reduced to a full-information feedback problem for an auxiliary plant, which is obtained from the given plant by solving an ($\mathcal{H}_\infty$ filtering) algebraic Riccati equation. Thus, the output-feedback problem can be reduced to a finite-dimensional convex programming problem over a set of real matrices. It is shown that the output-feedback controllers can always be chosen to have a structure similar to that of the standard $\mathcal{H}_\infty$ central controller. This implies that the order of (nearly) optimal output-feedback controllers need not exceed that of the generalized plant.

While the approach taken here is somewhat similar to the approach of Boyd and Barratt (1990), in that they also reduce such controller synthesis problems to convex optimization problems, there are significant differences between our results and those of Boyd and Barratt (1990). In particular, we reduce the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ controller synthesis problem to a convex optimization problem over a bounded subset of $q \times n$ and $n \times n$ symmetric real matrices, where q and n are, respectively the control input and the state dimensions. We accomplish this reduction of the problem without finite-dimensional approximations of the set of stabilizing controllers or frequency discretizations. Consequently, a solution to our convex programming problem is a global solution to the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis problem. This is considered to be an important contribution of our work. By comparison, the results of Boyd and Barratt (1990) applied to the present problem would reduce it to a convex optimization problem over the infinite-dimensional space of stable transfer functions.

Next, we briefly introduce notation used in this paper. The symbol $\varnothing$ denotes the empty set. Given a real matrix A, $\|A\|$ denotes its maximum singular value, $\operatorname{tr}(A)$ denotes its trace, and A' its transpose. We will say that a square matrix A is asymptotically stable if all its eigenvalues are inside the open unit disk. For A and B real symmetric matrices, $A > B$ (respectively $A \geq B$) iff the difference $A - B$ is positive-definite (respectively, positive-semi-definite). Linear time-invariant systems described by state space equations and are denoted by the script symbols, whereas the corresponding transfer matrices denoted by italics. For example, $\mathcal{G}$ denotes a system with transfer function G. The Hardy spaces $\mathcal{H}_2$ and $\mathcal{H}_\infty$ consist of matrix valued functions that are square integrable and essentially bounded, respectively, on the unit circle with analytic extension outside of the unit circle. The norms on these spaces are defined in the usual way.

2. THE MIXED $\mathcal{H}_2/\mathcal{H}_\infty$ PERFORMANCE MEASURE

In this section, we will define the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure. This will then be used in setting up the controller synthesis problem in the next section.

Let us begin by considering a finite-dimensional linear time-invariant discrete-time system $\mathcal{T}$ as shown in Fig. 2.

Suppose that $\mathcal{T}$ is internally stable with the discrete-time state-space model:

$$\begin{cases} (\sigma x)(k) := x(k+1) = Fx(k) + Gw(k), \\ z_0(k) = H_0 x(k) + J_0 w(k), \\ z_1(k) = H_1 x(k) + J_1 w(k), \end{cases} \tag{1}$$

where the matrices F, G, H_i and J_i are real and of compatible dimensions, and F has all eigenvalues in the open unit disk. (In the sequel, we will not show the time variable k explicitly in system equations.) Let

$$T_{zw} = \begin{bmatrix} T_{z_0w} \\ T_{z_1w} \end{bmatrix},$$

denote the transfer matrix from w to $z = [z_0', z_1']'$.

Let L_c denote the controllability gramian of the pair (F, G), i.e. L_c is the unique solution of the Lyapunov equation

$$FL_cF' + GG' = L_c. \tag{2}$$

Then, as is well known,

$$\|T_{z_0w}\|_2^2 = \operatorname{tr}(H_0L_cH_0' + J_0J_0').$$

Let $\gamma > 0$ be given, and consider the transfer matrix T_{z_1w}. In this paper, we will be interested in the $\mathcal{H}_\infty$ norm bound $\|T_{z_1w}\|_\infty < \gamma$. The

FIG. 2. Diagram for the definition of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure.

following theorem summarizes some useful results on characterizing this norm bound.

Theorem 2.1. Consider the internally stable linear time-invariant discrete-time system given by (1). Then the following statements are equivalent:

(1) $\|T_{z_1w}\|_\infty < \gamma$.

(2) There exists a nonsingular matrix P such that

$$\left\| \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} F & G \\ H_1/\gamma & J_1/\gamma \end{bmatrix} \begin{bmatrix} P^{-1} & 0 \\ 0 & I \end{bmatrix} \right\| < 1. \quad (3)$$

(3) There exists a real symmetric $Y > 0$ such that

$$\begin{bmatrix} F \\ H_1 \end{bmatrix} Y [F' \quad H_1'] + \begin{bmatrix} G \\ J_1 \end{bmatrix} [G' \quad J_1'] < \begin{bmatrix} Y & 0 \\ 0 & \gamma^2 I \end{bmatrix}. \quad (4)$$

Moreover, Y can be chosen to be the same as the one in item 4 below.

(4) There exists a real symmetric $Y > 0$ such that

$$M(Y) := \gamma^2 I - J_1 J_1' - H_1 Y H_1' > 0, \text{ and}$$
$$R(Y) := FYF' - Y + (FYH_1' + GJ_1')M^{-1} \times (H_1 YF' + J_1 G') + GG' < 0. \quad (5)$$

Moreover, Y can be chosen to be the same as the one in item 3 above.

(5) There exists a real symmetrical $Y \geq 0$ such that

$$M(Y) := \gamma^2 I - J_1 J_1' - H_1 Y H_1' > 0, \text{ and}$$
$$R(Y) := FYF' - Y + (FYH_1' + GJ_1')M^{-1} \times (H_1 YF' + J_1 G') + GG' = 0, \quad (6)$$

and $F + (FYH_1' + GJ_1')M^{-1}H_1$ is asymptotically stable. (In fact, Y satisfying the above conditions is unique.) Moreover, if $\hat{Y}$ denotes a solution to either (4) or (5), then $Y \leq \hat{Y}$.

Proof. The equivalence of items 1 and 5 can be found in Molinari (1975). The equivalence of items 1 and 4 follows from the equivalence of items 1 and 5 and a standard small perturbation argument. The equivalence of items 3 and 4 follows from simple algebraic manipulations and the Schur complement formula. Setting $Y := (P'P)^{-1}$ yields $2 \Rightarrow 3$, and setting $P = Y^{-1/2}$ gives $3 \Rightarrow 2$. Finally, in item 5, the inequality $Y \leq \hat{Y}$ follows from Ran and Vreugdenhil (1988). ∎

Now suppose $\|T_{z_1w}\|_\infty < \gamma$. Let Y denote the unique real symmetric matrix that satisfies condition 5 in Theorem 2.1. Then, from the definition of the controllability gramian L_c and Theorem 2.1, it follows that

$$0 \leq L_c \leq Y. \quad (7)$$

Note that this is the best possible upper bound for the controllability gramian that may be defined in terms of the solutions to the various quadratic matrix inequalities in Theorem 2.1.

Thus,

$$\|T_{z_0w}\|_2^2 = \operatorname{tr}(H_0 L_c H_0' + J_0 J_0') \leq \operatorname{tr}(H_0 Y H_0' + J_0 J_0').$$

The above inequality motivates the following definition of the *mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure* (*or cost*) $J(T_{zw})$ for the linear time-invariant system $\mathcal{T}$:

$$J(T_{zw}) := \operatorname{tr}(H_0 Y H_0' + J_0 J_0'). \quad (8)$$

The performance measure defined in (8) is the same as the one considered by Mustafa and Bernstein (1991), Bambang *et al.* (1990), and Haddad *et al.* (1991). (More precisely, this performance measure is one of the costs considered in Mustafa and Bernstein (1991).)

It is easily seen that $J(T_{zw})$ is only a function of the transfer matrix T_{zw}, and does not depend on the choice of realization, as long as such a realization is internally stable. This justifies our notation. Also $\|T_{z_0w}\|_2 \leq \sqrt{J(T_{zw})}$, and $\lim_{\gamma\to\infty} \sqrt{J(T_{zw})} = \|T_{z_0w}\|_2$. The mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure $J(T_{zw})$ is also a function of the parameter γ. In the sequel γ will remain fixed. Therefore, without loss of generality, we set

$$\gamma = 1,$$

for the remainder of this paper. Any other constraint level can be accommodated by simple scaling.

The following result provides an alternative characterization for the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure $J(T_{zw})$ that will be useful for establishing some of the results in this paper. The proof of this result is very similar to the proof of Lemma 1.1 in Khargonekar and Rotea (1991a). For the sake of brevity, details are omitted.

Lemma 2.2. Consider the stable system $\mathcal{T}$ defined in (1) and let T_{zw} denote the transfer matrix from w to z. Suppose that $\|T_{z_1w}\|_\infty < 1$. Let $M(.)$, $R(.)$ be given by (5), with $\gamma = 1$. Then

$$J(T_{zw}) = \inf\{\operatorname{tr}(H_0 Y H_0' + J_0 J_0') : Y = Y' > 0 \text{ such that } M(Y) > 0 \text{ and } R(Y) < 0\}.$$

3. THE MIXED $\mathcal{H}_2/\mathcal{H}_\infty$ CONTROL PROBLEM

In this section, we formulate the controller synthesis problem to be solved in this paper. Consider the finite-dimensional linear time-invariant discrete-time feedback system depicted in Fig. 3, where $\mathcal{G}$ is the generalized plant, including weighting functions, and $\mathcal{C}$ is the

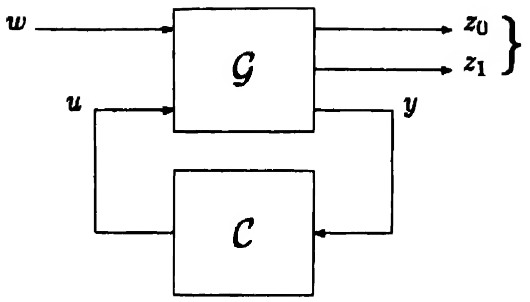


FIG 3. The synthesis framework.

controller to be designed. The signal w denotes an exogenous input, while z_0 and z_1 denote controlled (i.e. regulated) signals. The signals u and y denote the control input and the measured output, respectively. The transfer matrices of the plant and the controller are denoted by G and C, respectively. Let

$$T_{zw} = \begin{bmatrix} T_{z_0w} \\ T_{z_1w} \end{bmatrix},$$

denote the closed loop transfer matrix, where T_{z_0w} and T_{z_1w} are the closed loop transfer matrices from w to z_0 and w to z_1, respectively.

Definition 3.1. Let $\mathcal{G}$ and $\mathcal{C}$ be the given plant and controller. The controller $\mathcal{C}$ is called admissible (for the plant $\mathcal{G}$) if $\mathcal{C}$ internally stabilizes the plant $\mathcal{G}$. The set of all admissible controllers for the plant $\mathcal{G}$ is denoted by $\mathcal{A}(\mathcal{G})$. Furthermore, we define

$$\mathcal{A}_\infty(\mathcal{G}) := \{\mathcal{C} \in \mathcal{A}(\mathcal{G}) : \|T_{z_1w}\|_\infty < 1\}. \quad (9)$$

In the above notation, "$\mathcal{A}$" stands for "admissible". As is well known, $\mathcal{A}(\mathcal{G}) \neq \varnothing$ if and only if $\mathcal{G}$ is stabilizable from u and detectable from y. Also, the subscript "∞" in $\mathcal{A}_\infty(\mathcal{G})$ stands for the infinity norm constraint on T_{z_1w}.

Consider the feedback system shown in Fig. 3. Given a plant $\mathcal{G}$ and an internally stabilizing controller $\mathcal{C}$, the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ cost $J(T_{zw})$ of the closed loop system is a function of the transfer matrix T_{zw} only. We will denote this transfer matrix by $T_{zw}(G, C)$ and define

$$J(G, C) := J(T_{zw}(G, C)),$$

to emphasize on which plant and controller these closed loop quantities depend.

Following Khargonekar and Rotea (1991a), the sub-optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ controller synthesis problem considered in this paper is defined as follows.

The mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control problem. "Calculate the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure

$$v(\mathcal{G}) := \inf \{J(G, C) : \mathcal{C} \in \mathcal{A}_\infty(\mathcal{G})\}, \quad (10)$$

and, given any $\alpha > v(\mathcal{G})$, find a controller $\mathcal{C} \in \mathcal{A}_\infty(\mathcal{G})$ such that $J(G, C) < \alpha$".

In some cases involving state-feedback, it is natural to also consider memoryless, i.e. static controllers. In such a case, the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis problem is defined in the following way.

Definition 3.2. The set of static admissible controllers satisfying the $\mathcal{H}_\infty$ constraint is denoted by

$$\mathcal{A}_{\infty,m}(\mathcal{G}) := \{\mathcal{C} \in \mathcal{A}_\infty(\mathcal{G}) : C \in R^{q \times p}\}, \quad (11)$$

where $q = \dim(u)$ and $p = \dim(y)$.

The optimal performance over all admissible memoryless controllers is

$$v_m(\mathcal{G}) := \inf \{J(G, C) : \mathcal{C} \in \mathcal{A}_{\infty,m}(\mathcal{G})\}. \quad (12)$$

In (11) and (12), the subscript "m" stands for memoryless controllers.

The main results of this paper show that the computation of the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance (10), and the construction of a sub-optimal compensator, can be reduced to the convex optimization problem over a bounded convex subset of a space of real matrices.

4 STATE AND FULL-INFORMATION FEEDBACK PROBLEMS

In this section, we give a solution to the controller synthesis problem formulated previously, for the state/full-information feedback case. Here full-information feedback means that both the state and the exogenous inputs are available to the controller. Even though, such a feedback scheme is not realistic from a practical point of view, the full-information results are instrumental for addressing the more general case of output-feedback.

We will first ask the question whether the infimum of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure over all dynamic full-information feedback controllers equals the infimum over all static state-feedback controllers. In the continuous-time case, the answer to this question is in the affirmative (Khargonekar and Rotea, 1991a). However, *in the discrete-time*, the answer, in general, turns out to be in the negative. In fact, we will show that the infimum of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure over all dynamic full-information feedback controllers equals the infimum over all static full-information feedback controllers. Thus, the analogy with the continuous-time case breaks down in this sense. As mentioned in the Introduction, a similar situation also occurs in

the single objective standard $\mathcal{H}_\infty$ control problem in the discrete-time case.

On the other hand, if only the state of the system is available for feedback, then we will show that the infimum of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure over all dynamic state-feedback controllers equals the infimum over all static state-feedback controllers.

Finally, we will show that the static state/full-information controller synthesis problems may be effectively solved by means of finite-dimensional convex optimization.

Let us begin by considering the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis problem defined in Section 3 for the following plants (see also Fig. 3):

(1) State-feedback plant: the plant $\mathcal{G}$ is given by the state-space model

$$\mathcal{G}_{sf} := \begin{cases} \sigma x = Ax + B_1 w + B_2 u \\ z_0 = C_0 x + D_{01} w + D_{02} u \\ z_1 = C_1 x + D_{11} w + D_{12} u \\ y = x. \end{cases} \quad (13)$$

(2) Full-information plant: the plant $\mathcal{G}$ is given by the state-space model

$$\mathcal{G}_{fi} \begin{cases} \sigma x = Ax + B_1 w + B_2 u \\ z_0 = C_0 x + D_{01} w + D_{02} u \\ z_1 = C_1 x + D_{11} w + D_{12} u \\ y = [x' \quad w']'. \end{cases} \quad (14)$$

In the sequel, we let G_{sf} and G_{fi} denote the transfer matrices of (13) and (14), respectively. The subscripts "*sf*" and "*fi*" denote "state-feedback" and "full-information" structure, respectively. The only difference between the state-feedback and full-information plants is in the measurement equation. Note also that no assumptions on problem data, i.e. the matrices introduced in (13)–(14), are imposed.

4.1. *Reduction to memoryless feedback*

Theorem 4.1. Consider the full-information plant $\mathcal{G}_{fi}$ defined in (14). Then

$$\mathcal{A}_\infty(\mathcal{G}_{fi}) \neq \varnothing \Leftrightarrow \mathcal{A}_{\infty,m}(\mathcal{G}_{fi}) \neq \varnothing,$$

where $\mathcal{A}_\infty(\mathcal{G}_{fi})$ and $\mathcal{A}_{\infty,m}(\mathcal{G}_{fi})$ are as in (9) and (11), respectively. In this case,

$$\nu(\mathcal{G}_{fi}) = \nu_m(\mathcal{G}_{fi}),$$

where $\nu(\mathcal{G}_{fi})$ and $\nu_m(\mathcal{G}_{fi})$ denote the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ optimal costs in (10) and (12), respectively. Furthermore, given any $\alpha > \nu(\mathcal{G}_{fi})$, there exists a static full-information controller $\mathcal{K} \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi})$ such that $J(G_{fi}, K) < \alpha$.

Proof. We only need to show that if $\mathcal{A}_\infty(\mathcal{G}_{fi}) \neq \varnothing$, then $\mathcal{A}_{\infty,m}(\mathcal{G}_{fi}) \neq \varnothing$ and $\nu_m(\mathcal{G}_{fi}) \leq \nu(\mathcal{G}_{fi})$. Let $\epsilon > 0$ be given. From (10) it follows that there exists $\mathcal{C} \in \mathcal{A}_\infty(\mathcal{G}_{fi})$ such that

$$J(G_{fi}, C) \leq \nu(\mathcal{G}_{fi}) + \epsilon/2. \quad (15)$$

Let $\mathcal{C}$ be given by

$$\begin{cases} \sigma\xi = A_c\xi + B_{c1}x + B_{c2}w \\ u = C_c\xi + D_{c1}x + D_{c2}w. \end{cases} \quad (16)$$

The closed loop system corresponding to the interconnection of $\mathcal{G}_{fi}$ and $\mathcal{C}$ is given by:

$$\begin{cases} \sigma\eta = F\eta + Gw \\ z_0 = H_0\eta + J_0 w \\ z_1 = H_1\eta + J_1 w, \end{cases}$$

where

$$F := \begin{bmatrix} A + B_2 D_{c1} & B_2 C_c \\ B_{c1} & A_c \end{bmatrix},$$

$$G := \begin{bmatrix} B_1 + B_2 D_{c2} \\ B_{c2} \end{bmatrix},$$

$$H_0 := [C_0 + D_{02}D_{c1} \quad D_{02}C_c],$$

$$J_0 := D_{01} + D_{02}D_{c2},$$

$$H_1 := [C_1 + D_{12}D_{c1} \quad D_{12}C_c],$$

$$J_1 := D_{11} + D_{12}D_{c2}.$$

Since $\mathcal{C} \in \mathcal{A}_\infty(\mathcal{G}_{fi})$, it follows that F is stable and $\|T_{z_1w}(G_{fi}, C)\|_\infty < 1$. Using Theorem 2.1 and Lemma 2.2 we may now conclude that $\exists Y = Y' > 0$ such that:

$$\begin{aligned} &(1)\ M := I - J_1J_1' - H_1YH_1' > 0 \\ &(2)\ R(Y) := FYF' - Y + (FYH_1' + GJ_1') \\ &\qquad \times M^{-1}(H_1YF' + J_1G') + GG' < 0, \\ &(3)\ \mathrm{tr}\,(H_0YH_0' + J_0J_0') \leq J(G_{fi}, C) + \epsilon/2. \end{aligned} \quad (17)$$

Now combining (15), and item 3 in (17), we get

$$\mathrm{tr}\,(H_0YH_0' + J_0J_0') \leq \nu(G_{fi}) + \epsilon. \quad (18)$$

Using the matrix Y introduced in (17), we will construct a memoryless controller $\mathcal{K}$ for $\mathcal{G}_{fi}$. Let $n = \dim(x)$ and $n_c = \dim(\xi)$ and partition Y and $R(Y)$ according to the plant and controller dimensions, i.e.

$$Y = \begin{bmatrix} Y_1 & Y_2 \\ Y_2' & Y_3 \end{bmatrix}, \quad R(Y) = \begin{bmatrix} R_1 & R_2 \\ R_2' & R_3 \end{bmatrix},$$

where $\dim(Y_1) = n \times n$, $\dim(Y_2) = n \times n_c$, $\dim(Y_3) = n_c \times n_c$, and similarly for $R(Y)$. Note that $Y_1 > 0$ and $R_1 < 0$. Define the memoryless full-information controller $\mathcal{K}$ by

$$u = K_1x + K_2w,$$

where

$$K_1 := D_{c1} + C_cY_2'Y_1^{-1}, \quad K_2 := D_{c2}. \quad (19)$$

The closed loop system resulting from the interconnection of $\mathcal{G}_{fi}$ and $\mathcal{K}$ is given by

$$\begin{cases} \sigma x = F_m x + G_m w \\ z_0 = H_{0m} x + J_{0m} w \\ z_1 = H_{1m} x + J_{1m} w, \end{cases}$$

where $F_m := A + B_2K_1$, $G_m := B_1 + B_2K_2$, $H_{0m} := C_0 + D_{02}K_1$, $J_{0m} := D_{01} + D_{02}K_2$, $H_{1m} := C_1 + D_{12}K_1$, $J_{1m} := D_{11} + D_{12}K_2$, and the gains K_1 and K_2 are given by (19).

Define $Q := Y_3 - Y_2'Y_1^{-1}Y_2$. Since, $Y > 0$ it follows that $Q > 0$ by taking Schur complement. Simple algebraic manipulations show that the "1–1 block" of $R(Y)$ satisfies

$$\begin{aligned} R_1 = F_mY_1F_m' - Y_1 + G_mG_m' + B_2C_cQC_c'B_2' \\ + (F_mY_1H_{1m}' + G_mJ_{1m}' + B_2C_cQC_c'D_{12}')M^{-1} \\ \times (H_{1m}Y_1F_m' + J_{1m}B_{1m}' + D_{12}C_cQC_c'B_2') < 0, \end{aligned} \tag{20}$$

where

$$\begin{aligned} M = I - J_{1m}J_{1m}' - H_{1m}Y_1H_{1m}' \\ - D_{12}C_cQC_c'D_{12}' > 0. \end{aligned} \tag{21}$$

Since $Y_1 > 0$, (20), (21) and the implication $4 \Rightarrow 3$ of Theorem 2.1 imply that Y_1 satisfies

$$\begin{bmatrix} F_m \\ H_{1m} \end{bmatrix} Y_1 [F_m' \quad H_{1m}'] + \begin{bmatrix} G_m \\ J_{1m} \end{bmatrix} [G_m' \quad J_{1m}'] + \begin{bmatrix} B_2 \\ D_{12} \end{bmatrix} C_cQC_c' [B_2' \quad D_{12}'] < \begin{bmatrix} Y_1 & 0 \\ 0 & I \end{bmatrix}.$$

It follows that

$$\begin{bmatrix} F_m \\ H_{1m} \end{bmatrix} Y_1 [F_m' \quad H_{1m}'] + \begin{bmatrix} G_m \\ J_{1m} \end{bmatrix} [G_m' \quad J_{1m}'] < \begin{bmatrix} Y_1 & 0 \\ 0 & I \end{bmatrix}, \tag{22}$$

since $Q > 0$.

Now using (22) and $Y_1 > 0$, a simple Lyapunov argument shows that F_m is a stable matrix Further, from implication $3 \Rightarrow 1$ in Theorem 2.1, it follows that $\|T_{z_1w}(G_{fi}, K)\|_\infty < 1$. Consequently, the memoryless controller defined in (19) satisfies $\mathcal{K} \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi})$.

It is easy to verify that

$$\begin{aligned} H_{0m}Y_1H_{0m}' + J_{0m}J_{0m}' \\ = H_0YH_0' + J_0J_0' - D_{02}C_cQC_c'D_{02}'. \end{aligned}$$

Since $Q > 0$, we may now conclude that

$$\mathrm{tr}\,(H_{0m}Y_1H_{0m}' + J_{0m}J_{0m}') \le \mathrm{tr}\,(H_0YH_0' + J_0J_0'),$$

which, together with (22), implies

$$\mathrm{tr}\,(H_{0m}Y_1H_{0m}' + J_{0m}J_{0m}') \le \nu(\mathcal{G}_{fi}) + \epsilon.$$

Using the last inequality, $Y_1 > 0$, the implication $3 \Rightarrow 4$ in Theorem 2.1, Lemma 2.2, and (18) we obtain

$$\begin{aligned} \nu_m(\mathcal{G}_{fi}) \le J(G_{fi}, K) \\ \le \mathrm{tr}\,(H_{0m}Y_1H_{0m}' + J_{0m}J_{0m}') \le \nu(\mathcal{G}_{fi}) + \epsilon. \end{aligned}$$

Since ϵ is arbitrary, we conclude that

$$\nu_m(\mathcal{G}_{fi}) \le \nu(\mathcal{G}_{fi}).$$

The last part of the theorem now follows from definitions. ■

Theorem 4.1 applies in the case where both the state and the exogenous inputs are available for feedback. This is rarely, if every, true in practice. The principal motivation for this result comes from the output-feedback case. As will be seen in the next section, Theorem 4.1 is a key result in the derivation of the output feedback solution.

However, in applications one often comes across problems where the state vector is available for feedback. From this point of view the following results on the state-feedback case is much more useful.

Theorem 4.2. Consider the state-feedback plant $\mathcal{G}_{fi}$ defined in (13). Then

$$\mathcal{A}_\infty(\mathcal{G}_{sf}) \ne \emptyset \Leftrightarrow \mathcal{A}_{\infty,m}(\mathcal{G}_{sf}) \ne \emptyset,$$

where $\mathcal{A}_\infty(\mathcal{G}_{sf})$ and $\mathcal{A}_{\infty,m}(\mathcal{G}_{sf})$ are as in (9) and (11), respectively. In this case

$$\nu(\mathcal{G}_{sf}) = \nu_m(\mathcal{G}_{sf}),$$

where $\nu(\mathcal{G}_{sf})$ and $\nu_m(\mathcal{G}_{sf})$ denote the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ optimal costs defined in (10) and (12), respectively. Furthermore, given any $\alpha > \nu(\mathcal{G}_{sf})$, there exists $\mathcal{K} \in \mathcal{A}_{\infty,m}(\mathcal{G}_{sf})$, such that $J(G_{sf}, K) < \alpha$.

A proof of this result can be constructed from the proof of Theorem 4.1 by setting B_{c2} and D_{c2} equal to zero in the definition of the controller $\mathcal{C}$ given in (16). Details are omitted for the sake of brevity.

4.2. *A convex optimization approach to static feedback problem*

In this subsection we will show that the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem with state/full-information feedback can be reduced to a convex optimization problem over a convex bounded set of real matrices. That is, given $\alpha > \nu_m(\mathcal{G}_{fi})$, real matrices K_1 and K_2 such that $J(G_{fi}, [K_1 \quad K_2]) < \alpha$ can be found by solving a finite-dimensional convex programming problem. We will consider the full-information case. *The state-feedback case follows by taking $K_2 = 0$ in the analysis below.*

With reference to the full-information plant

defined in (14), let $n = \dim(x)$, $q = \dim(u)$, $p = \dim(w)$. Let Σ denote the set of all real $n \times n$ symmetric matrices, and define

$$\Omega := \{(W, Y, K_2) \in R^{q\times n} \times \Sigma \times R^{q\times p} : Y > 0\}. \tag{23}$$

Note that Ω is a strictly convex open subset of $R^{q\times n} \times \Sigma \times R^{q\times p}$. Given $(W, Y, K_2) \in \Omega$ define

$$\begin{aligned} f(W, Y, K_2) \\ := \operatorname{tr}((C_0Y + D_{02}W)Y^{-1}(C_0Y + D_{02}W)' \\ + (D_{01} + D_{02}K_2)(D_{01} + D_{02}K_2)'). \end{aligned} \tag{24}$$

Given any $(W, Y, K_2) \in \Omega$, define

$$\begin{aligned} L(W, Y, K_2) \\ := \begin{bmatrix} AY + B_2W \\ C_1Y + D_{12}W \end{bmatrix} Y^{-1} \begin{bmatrix} AY + B_2W \\ C_1Y + D_{12}W \end{bmatrix}' \\ + \begin{bmatrix} B_1 + B_2K_2 \\ D_{11} + D_{12}K_2 \end{bmatrix} \begin{bmatrix} B_1 + B_2K_2 \\ D_{11} + D_{12}K_2 \end{bmatrix}' - \begin{bmatrix} Y & 0 \\ 0 & I \end{bmatrix}. \end{aligned} \tag{25}$$

Now consider the set of real matrices:

$$\Phi(\mathcal{G}_{fi}) := \{(W, Y, K_2) \in \Omega : L(W, Y, K_2) < 0\}, \tag{26}$$

and the constrained optimization problem

$$\psi(\mathcal{G}_{fi}) := \inf\{f(W, Y, K_2) : (W, Y, K_2) \in \Phi(\mathcal{G}_{fi})\}. \tag{27}$$

Note that $\psi(\mathcal{G}_{fi}) \geq 0$ since $f \geq 0$ on Ω, as can be seen from (24). We now state the main result of this subsection.

Theorem 4.3. Consider the system $\mathcal{G}_{fi}$ defined in (14) with transfer matrix G_{fi}. Let $\mathcal{A}_{\infty,m}(\mathcal{G}_{fi})$ be the set of static controllers defined in (11). Then,

$$\mathcal{A}_{\infty,m}(\mathcal{G}_{fi}) \neq \varnothing \Leftrightarrow \Phi(\mathcal{G}_{fi}) \neq \varnothing,$$

where $\Phi(\mathcal{G}_{fi})$ is given by (26). In this case

$$\nu_m(\mathcal{G}_{fi}) = \psi(\mathcal{G}_{fi}),$$

where $\nu_m(\mathcal{G}_{fi})$ and $\psi(\mathcal{G}_{fi})$ are as in (12) and (27), respectively. Furthermore, given any $\alpha > \nu_m(\mathcal{G}_{fi})$, there exists a triple $(W, Y, K_2) \in \Phi(\mathcal{G}_{fi})$ such that the static full-information controller

$$K := [WY^{-1} \quad K_2],$$

satisfies

$$\mathcal{K} \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi}) \quad \text{and} \quad J(G_{fi}, K) < \alpha.$$

This result is a direct and straightforward generalization of Theorem 4.2 of Khargonekar and Rotea (1991a). Proof is omitted.

In the remainder of this section we will show that the optimization problem defined in (27) is convex.

Lemma 4.4. Let Ω denote the set defined in (23). The mapping $f : \Omega \rightarrow R^+$ defined in (24) is a real-analytic convex function on Ω.

Proof. Using (24) f may be rewritten as

$$\begin{aligned} f(W, Y, K_2) = \operatorname{tr}(C_0YC_0') + 2\operatorname{tr}(C_0'D_{02}W) \\ + \operatorname{tr}(D_{02}WY^{-1}W'D_{02}') \\ + \operatorname{tr}(D_{01} + D_{02}K_2)(D_{01}' + K_2'D_{02}'). \end{aligned}$$

In Khargonekar and Rotea (1991a) it was shown that the first three terms in this expression are convex in Ω. Clearly, the term $\operatorname{tr}(D_{01} + D_{02}K_2)(D_{01}' + K_2'D_{02}')$ is convex in K_2. The convexity of $f(.)$ now follows. The fact that $f(.)$ is real analytic follows from its definition. ■

Lemma 4.5. Let $L : \Omega \rightarrow \Sigma$ denote the matrix-valued mapping defined in (25). Then L is a convex mapping. Consequently, the constraint set $\Phi(\mathcal{G}_{fi})$ defined in (26) is convex.

Proof. Let $(W, Y, K_2) \in \Omega$. Using (25), it is easy to see that $L(W, Y, K_2)$ can be rewritten as

$$\begin{aligned} L(W, Y, K_2) = \begin{bmatrix} A & B_2 \\ C_1 & D_{12} \end{bmatrix} \begin{bmatrix} Y \\ W \end{bmatrix} Y^{-1} [Y' \quad W'] \\ \times \begin{bmatrix} A' & C_1' \\ B_2' & D_{12}' \end{bmatrix} + \begin{bmatrix} B_1 & B_2 \\ D_{11} & D_{12} \end{bmatrix} \\ \times \begin{bmatrix} I \\ K_2 \end{bmatrix} [I \quad K_2'] \begin{bmatrix} B_1' & D_{11}' \\ B_2' & D_{12}' \end{bmatrix} \\ - \begin{bmatrix} Y & 0 \\ 0 & I \end{bmatrix} \\ = \bar{F}\bar{W}Y^{-1}\bar{W}'\bar{F}' + \bar{G}\bar{K}\bar{K}'\bar{G}' \\ - \begin{bmatrix} Y & 0 \\ 0 & I \end{bmatrix}, \end{aligned}$$

where

$$\bar{F} := \begin{bmatrix} A' & C_1' \\ B_2' & D_{12}' \end{bmatrix}, \quad \bar{G} := \begin{bmatrix} B_1 & B_2 \\ D_{11} & D_{12} \end{bmatrix},$$

$$\bar{W} := [Y' \quad W']', \qquad \bar{K} := [I \quad K_2']'.$$

By Proposition E.7.f in Marshall and Olkin (1979) the mappings $(\bar{W}, Y) \rightarrow \bar{F}\bar{W}Y^{-1}\bar{W}'\bar{F}'$, and $\bar{K} \rightarrow \bar{G}\bar{K}\bar{K}'\bar{G}'$, are convex on their domains (here $Y = Y' > 0$). Since the maps $(W, Y) \rightarrow [Y' \ W']'$ and $K_2 \rightarrow [I \ K_2']'$ are affine linear, the convexity of L follows. Finally, the convexity of $\Phi(\mathcal{G}_{fi})$ follows from the convexity of L. ■

Lemma 4.6. Consider the set $\Phi(\mathcal{G}_{fi})$ defined in (26). Assume that D_{12} has full column rank. Suppose that for all z inside the open unit disc, the system matrix

$$\begin{bmatrix} zI - A & B_2 \\ -C_1 & D_{12} \end{bmatrix},$$

has full column rank. Then the set $\Phi(\mathcal{G}_{fi})$ is bounded.

Proof. We need to show that there exist positive constants $m_1, m_2, m_3 < \infty$ such that

$$(W, Y, K_2) \in \Phi(\mathcal{G}_{fi}) \Rightarrow \|W\| \le m_1,$$
$$\|Y\| \le m_2, \quad \|K_2\| \le m_3.$$

Let $(W, Y, K_2) \in \Phi(\mathcal{G}_{fi})$. From definitions of $\Phi(\mathcal{G}_{fi})$ and $L(W, Y, K_2)$, it follows that $D_{12}K_2K_2'D_{12}' < I$, which implies that $\|D_{12}K_2\| < 1$. Since D_{12} is full column rank there exists a constant $m_3 < \infty$, such that

$$\|K_2\| \le m_3 < \infty.$$

Now define the matrices

$$T: \begin{bmatrix} I & -B_2XD_{12}' \\ 0 & I - D_{12}XD_{12}' \end{bmatrix},$$
$$\bar{C}_1 := (I - D_{12}XD_{12}')C_1,$$
$$\bar{A} := A - B_2XD_{12}'C_1,$$

where $X := (D_{12}'D_{12})^{-1}$. Premultiply L by T and postmultiply it by T'. After simple algebraic manipulations we obtain

$$\begin{bmatrix} \bar{A} \\ \bar{C}_1 \end{bmatrix} Y [\bar{A}' \quad \bar{C}_1'] - \begin{bmatrix} Y & 0 \\ 0 & I - D_{12}XD_{12}' \end{bmatrix} \le 0. \quad (28)$$

To prove the boundedness of Y, we are going to use Theorem 2.1 of Section 2 and Theorem 3.1 in Ran and Vreugdenhil (1988). To be able to use these theorems, we first need to show that $I - \bar{C}_1Y\bar{C}_1' > 0$. It follows from (28) that $\bar{C}_1Y\bar{C}_1' + D_{12}XD_{12}' - I \le 0$, and so $I - \bar{C}_1Y\bar{C}_1' \ge 0$. Now suppose there exists a vector $y \ne 0 \in R^p$ such that

$$(I - \bar{C}_1Y\bar{C}_1')y = 0. \quad (29)$$

Multiplying the inequality $\bar{C}_1Y\bar{C}_1' + D_{12}XD_{12}' - I \le 0$ by y on the right and y' on the left, we get $y'D_{12}XD_{12}'y = 0$. Since X is invertible, it follows that

$$D_{12}'y = 0 \quad \text{and} \quad \bar{C}_1'y = C_1'y. \quad (30)$$

Also by taking the $(2, 2)$ sub-block of the inequality $L < 0$, we get

$$I - (C_1Y + D_{12}W)Y^{-1}(C_1Y + D_{12}W)' - D_{12}K_2K_2'D_{12}' > 0. \quad (31)$$

Premultiplying (31) by y' and postmultiplying it by y yields

$$y'(I - C_1YC_1')y > 0.$$

Since $C_1'y = \bar{C}_1'y$, it now follows that

$$0 = y'(I - \bar{C}_1Y\bar{C}_1')y = y'(I - C_1YC_1')y > 0.$$

This contradicts (29), and therefore

$$\bar{X} := I - \bar{C}_1Y\bar{C}_1' > 0. \quad (32)$$

Using (32) and taking Schur complement of the $(2, 2)$ block in (28), it follows that

$$\bar{R} := \bar{A}Y\bar{A}' - Y - \bar{A}Y\bar{C}_1'\bar{X}^{-1}\bar{C}_1Y\bar{A}' \le 0. \quad (33)$$

Now the rank assumption on the system matrix ensures that the pair $(\bar{C}_1, \bar{A})$ has no unobservable modes inside the unit disc. Using a simple extension of Theorem 3.1 in Ran and Vreugdenhil (1988) and the observability of the stable modes of $(\bar{C}_1, \bar{A})$, it follows that there exists a real symmetric matrix Y_- (depending only on $\bar{A}$, B and $\bar{C}_1$) such that $-Y \ge Y_-$, or $Y \le -Y_-$. Let $m_1 = \|Y_-\|$, then since $Y > 0$, $\|Y\| \le m_1 < \infty$.

Finally, from equation (31) it follows that

$$(C_1Y + D_{12}W)Y^{-1}(C_1Y + D_{12}W)' < I$$
$$\Rightarrow \|(C_1Y + D_{12}W)Y^{-1/2}\| < 1$$
$$\Rightarrow Y^{-1/2}(C_1Y + D_{12}W)(C_1Y + D_{12}W)'Y^{-1/2} < I$$
$$\Rightarrow \|C_1Y + D_{12}W\| < \sqrt{\|Y\|}.$$

Since $\|Y\| \le m_1$ and D_{12} has full column rank, we conclude that there exists $m_2 < \infty$ such that

$$\|W\| \le m_2 < \infty. \qquad \blacksquare$$

5. OUTPUT-FEEDBACK CASE

In this section we will solve the synthesis problem defined in Section 3 for the output-feedback case. We will show that the problem can be reduced to solving one algebraic Riccati equation, and a convex optimization problem similar to the one in Section 4. In the following subsection we introduce a technical result that will be needed in order to prove the main theorem of this section.

5.1. *Preliminaries*

The next result is an extension of Redheffer's lemma (see, for example, Iglesias and Glover (1991) and Stoorvogel (1990)) to the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure for discrete-time systems.

Consider the feedback interconnection of Fig.

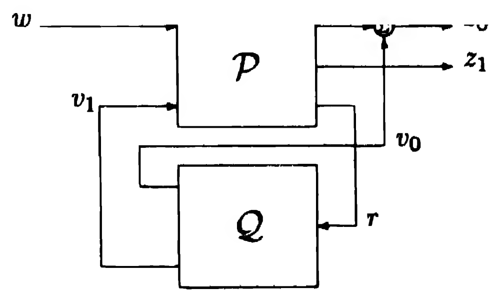


FIG. 4. System interconnection in Lemma 5.1.

4, where

$$\mathcal{P} := \begin{cases} \sigma\eta = A\eta + B_1 w + B_2 v_1 \\ \bar{z}_0 = C_0\eta \\ z_0 = \bar{z}_0 + v_0 \\ z_1 = C_1\eta + D_{11}w + D_{12}v_1 \\ r = C_2\eta + D_{21}w + D_{22}v_1, \end{cases} \quad (34)$$

and

$$\mathcal{Q} := \begin{cases} \sigma x = \hat{A}x + \hat{B}r \\ v_0 = \hat{C}_0 x + \hat{D}_0 r \\ v_1 = \hat{C}_1 x + \hat{D}_1 r. \end{cases} \quad (35)$$

The matrices in the state-space equations (34) and (35) are real and of compatible dimensions. Let P and Q denote the corresponding transfer matrices, and partition them as

$$P = \begin{array}{c} \bar{z}_0 \{ \\ z_1, r \{ \end{array} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix}, \quad Q = \begin{array}{c} v_0 \{ \\ v_1 \{ \end{array} \begin{bmatrix} Q_0 \\ Q_1 \end{bmatrix}. \quad (36)$$

Lemma 5.1. Consider the feedback system shown in Fig. 4, where $\mathcal{P}$ and $\mathcal{Q}$ are given by (34) and (35), respectively. Let T_{zw} denote the closed loop transfer matrix from w to $z = (z_0, z_1)$. Suppose that $\mathcal{P}$ is internally stable, and let L_c denote the controllability gramian of the pair $(A, [B_1 \;\; B_2])$, i.e.

$$AL_cA' + B_1B_1' + B_2B_2' = L_c. \quad (37)$$

Suppose also that D_{12} is square and nonsingular, that $A - B_2D_{12}^{-1}C_1$ is a stable matrix, and

$$[B_1 \;\; B_2]\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}' + AL_c[C_1' \;\; C_2'] = 0, \quad (38)$$

$$\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}' + \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} L_c[C_1' \;\; C_2'] = I. \quad (39)$$

Then the following statements are equivalent:

(i) The feedback system in Fig. 4 is well-posed, internally stable, and $\|T_{z_1w}\|_\infty < 1$.

(ii) $\mathcal{Q}$ is internally stable and $\|Q_1\|_\infty < 1$.

If the above conditions hold, the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ costs $J(T_{zw})$ (with respect to z_0) and $J(Q)$ (with respect to v_0), are defined and they satisfy

$$J(T_{zw}) = \mathrm{tr}\,(C_0L_cC_0') + 2\mathrm{tr}\,(\hat{D}_0C_2L_cC_0') + J(\mathcal{Q}). \quad (40)$$

A continuous-time version of the above result is in Khargonekar and Rotea (1991a). The discrete-time case differs from the continuous-time case in one important aspect. In the continuous-time case the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ cost $J(T_{zw})$ of the feedback interconnection of $\mathcal{P}$ and $\mathcal{Q}$, is the sum of the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ cost of $\mathcal{Q}$ and the $\mathcal{H}_2$ norm of the "1–1" block of the inner plant $\mathcal{P}$. This decomposition fails in the discrete-time case; there is an additional cross term (which depends on $\mathcal{Q}$) in (40).

Proof. The proof of the equivalence of the statements (i) and (ii) can be found in Doyle *et al.* (1989b); Iglesias and Glover (1991) and Stoorvogel (1990).

Suppose now that either one of these statements is true. For simplicity, the rest of the proof will be done under the assumption that P_{22} is strictly proper, i.e. $D_{22} = 0$. Let $\psi = [x' \;\; \eta']'$ denote the state of the composite system. It is easy to check that the system resulting from the interconnection of $\mathcal{P}$ and $\mathcal{Q}$ is given by

$$\begin{cases} \sigma\psi = F\psi + Gw \\ z_0 = H_0\psi + J_0w \\ z_1 = H_1\psi + J_1w, \end{cases} \quad (41)$$

where

$$F := \begin{bmatrix} \hat{A} & \hat{B}C_2 \\ B_2\hat{C}_1 & A + B_2\hat{D}_1C_2 \end{bmatrix},$$

$$G := \begin{bmatrix} \hat{B}D_{21} \\ B_1 + B_2\hat{D}_1D_{21} \end{bmatrix},$$

$$H_0 := [\hat{C}_0, \; C_0 + \hat{D}_0C_2],$$

$$H_1 := [D_{12}\hat{C}_1 \quad C_1 + D_{12}\hat{D}_1C_2],$$

$$J_0 := \hat{D}_0D_{21}, \quad J_1 := D_{11} + D_{12}\hat{D}_1D_{21}.$$

Note that because of internal stability, all eigenvalues of F are inside the open unit disc. Moreover, $\|T_{z_1w}\|_\infty < 1$.

In order to establish formula (40), we need to determine the stabilizing solution of the ARE corresponding to the condition $\|T_{z_1w}\|_\infty < 1$. That is, the real symmetric matrix Y such that

$$M(Y) := I - J_1J_1' - H_1YH_1' > 0,$$

$$R(Y) := FYF' - Y + (FYH_1' + GJ_1')M^{-1} \times (H_1YF' + J_1G') + GG' = 0, \quad (42)$$

and $F + (FYH_1' + GJ_1')M^{-1}H_1$ is asymptotically stable. Since $\hat{A}$ is asymptotically stable $\|Q_1\|_\infty < 1$, there exists a real symmetric matrix $\hat{Y}$ such that

$$\hat{M}(\hat{Y}) := I - \hat{D}_1\hat{D}_1' - \hat{C}_1\hat{Y}\hat{C}_1' > 0,$$

$$\hat{R}(\hat{Y}) := \hat{A}\hat{Y}\hat{A}' - \hat{Y} + (\hat{A}\hat{Y}\hat{C}_1' + \hat{B}\hat{D}_1')\hat{M}^{-1} \times (\hat{C}_1\hat{Y}\hat{A}' + \hat{D}_1\hat{B}') + \hat{B}\hat{B}' = 0, \quad (43)$$

and $\hat{A} + (\hat{A}\hat{Y}\hat{C}_1' + \hat{B}\hat{D}_1')\hat{M}^{-1}\hat{C}_1$ stable.

Let L_c be given by (37) and set

$$Y := \begin{bmatrix} \hat{Y} & 0 \\ 0 & L_c \end{bmatrix}. \quad (44)$$

After some algebra and using equations (43) and (37)–(39), we obtain $M(Y) = D_{12}\hat{M}(\hat{Y})D_{12}' > 0$

and $R(Y)=0$. Moreover,

$$F+(FYH_1'+GJ_1')M^{-1}H_1 = \begin{bmatrix} \hat{A}+(\hat{A}\hat{Y}\hat{C}_1'+\hat{B}\hat{D}_1')\hat{M}^{-1}\hat{C}_1 & * \\ 0 & -B_2D_{12}^{-1}C_1 \end{bmatrix},$$

where the "* block" is not relevant. Since $\hat{A}+(\hat{A}\hat{Y}\hat{C}_1'+\hat{B}\hat{D}_1')\hat{M}^{-1}\hat{C}_1$ and $A-B_2D_{12}^{-1}C_1$ are both asymptotically stable, we conclude that the real symmetric matrix Y defined in (44) is the unique stabilizing solution of the ARE (42).

Next we compute the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ cost $J(T_{zw})$. From equations (39), (41) and (44), it follows that

$$\begin{aligned} J(T_{zw}) &= \operatorname{tr}(H_0YH_0'+J_0J_0') = \operatorname{tr}(C_0L_cC_0') \\ &\quad + \operatorname{tr}(\hat{C}_0\hat{Y}\hat{C}_0'+\hat{D}_0\hat{D}_0') \\ &\quad + \operatorname{tr}(\hat{D}_0C_2L_cC_0'+C_0L_cC_2'\hat{D}_0') \\ &= \operatorname{tr}(C_0L_cC_0')+J(Q) \\ &\quad + 2\operatorname{tr}(\hat{D}_0C_2L_cC_0'). \quad \blacksquare \end{aligned} \tag{45}$$

5.2. *Main result*

We now consider the output-feedback mixed $\mathcal{H}_2/\mathcal{H}_\infty$ controller synthesis problem. Suppose the plant $\mathcal{G}$ in Fig. 3 is given by the state-space model:

$$\mathcal{G}:=\begin{cases} \sigma x = Ax+B_1w+B_2u \\ z_0 = C_0x+D_0u \\ z_1 = C_1x+D_1u \\ y = C_2x+D_2w, \end{cases} \tag{46}$$

where all the matrices in (46) are constant real matrices of compatible dimensions. We will also make the following assumptions:

(A1) The triple (C_2, A, B_2) is stabilizable and detectable.

(A2) For each complex number z, such that $|z|=1$, the matrix

$$\begin{bmatrix} zI-A & -B_1 \\ C_2 & D_2 \end{bmatrix},$$

has full row rank.

Note that we have also assumed that there are no feedthrough terms from w to z_i, or u to y. Even though it is possible to include these terms, we have chosen not to do so, to keep the presentation as simple as possible. The results given below may be combined with those in Stoorvogel (1990) to obtain formulae for the most general case.

Suppose there exists admissible controller $\mathscr{C}$ such that the closed loop system is internally stable and $\|T_{z_1w}\|_\infty<1$, i.e. $\mathscr{C}\in\mathscr{A}_\infty(\mathcal{G})$. Then it follows (Stoorvogel, 1990) that there exists a (unique) real symmetric matrix $Q\geq 0$ such that

$$V:=C_2QC_2'+D_2D_2'>0, \tag{47}$$

$$R:=I-C_1QC_1'+C_1QC_2'V^{-1}C_2QC_1'>0, \tag{48}$$

and Q satisfies the following discrete-time algebraic Riccati equation:

$$AQA'-Q+B_1B_1'-[AQC_2'+B_1D_2' \quad AQC_1'] \times G(Q)^{-1}\begin{bmatrix} C_2QA'+D_2B_1' \\ C_1QA' \end{bmatrix}=0, \tag{49}$$

where

$$G(Q):=\begin{bmatrix} D_2D_2' & 0 \\ 0 & -I \end{bmatrix}+\begin{bmatrix} C_2 \\ C_1 \end{bmatrix}Q[C_2' \quad C_1'].$$

Moreover, the matrix

$$A-[AQC_2'+B_1D_2' \quad AQC_1']G(Q)^{-1}\begin{bmatrix} C_2 \\ C_1 \end{bmatrix}, \tag{50}$$

is asymptotically stable.

Given a real symmetric matrix Q, define the auxiliary full-information system:

$$\mathcal{G}_{fi}(Q):=\begin{cases} \sigma x_g = (A+ZR^{-1}C_1)x_g \\ \qquad + (AQC_2'+B_1D_2' \\ \qquad + ZR^{-1}C_1QC_2')V^{-1/2}r \\ \qquad + (B_2+ZR^{-1}D_1)u \\ \quad =: A_gx_g+B_{1g}r+B_{2g}u, \\ \hat{v}_0 = C_0x_g+C_0QC_2'V^{-1/2}r+D_0u \\ \quad =: C_{0g}x_g+D_{01g}r+D_{02g}u \\ v_1 = R^{-1/2}C_1x_g \\ \qquad + R^{-1/2}C_1QC_2'V^{-1/2}r \\ \qquad + R^{-1/2}D_1u \\ \quad =: C_{1g}x_g+D_{11g}r+D_{12g}u \\ y = [x_g' \quad w']', \end{cases} \tag{51}$$

where $Z:=AQC_1'-(AQC_2'+B_1D_2')V^{-1}C_2QC_1'$, and A, C_i, D_i, B_2 are given in (46). Let $G_{fi}(Q)$ denote the transfer matrix from (w, u) to $(\hat{v}_0, v_1, v)$ in (51). The next result gives the solution to the mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problem for the case of output feedback.

Theorem 5.2. Consider the plant $\mathcal{G}$ defined in (46). Suppose that Assumptions A1–A2 hold. Let $\mathscr{A}_\infty(\mathcal{G})$ be as defined in (9), and suppose that $\mathscr{A}_\infty(\mathcal{G})\neq\varnothing$. Then there exists a unique (real symmetric) matrix $Q\geq 0$ that satisfies the conditions (47)–(50). Let $\mathcal{G}_{fi}(Q)$ denote the auxiliary system defined in (51). Then the following statements hold:

(1) The set of admissible controllers $\mathscr{A}_\infty(\mathcal{G}_{fi}(Q))$ is nonempty, and the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance measure is given by

$$\begin{aligned} v(\mathcal{G}) &= \operatorname{tr}(C_0QC_0') \\ &\quad - \operatorname{tr}(C_0QC_2'V^{-1}C_2QC_0')+v(\mathcal{G}_{fi}(Q)), \end{aligned} \tag{52}$$

where $v(\mathcal{G}_{fi}(Q))$ is the optimal mixed $\mathcal{H}_2/\mathcal{H}_\infty$ performance for the auxiliary full-information plant $\mathcal{G}_{fi}(Q)$.
(2) Given any $\alpha > v(\mathcal{G})$, there exists a static full information controller

$$\mathcal{K} := K_1 x_g + K_2 w, \tag{53}$$

such that $\mathcal{K} \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi}(Q))$ and

$$J(G_{fi}(Q), K) < \alpha - \mathrm{tr}\,(C_0 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0').$$

(3) For any full-information controller $\mathcal{K} = [\mathcal{K}_1 \mathcal{K}_2] \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi}(Q))$, the dynamic output-feedback controller

$$\mathscr{C} := \begin{cases} \sigma\xi = F\xi + Gy \\ u = H\xi + Jy, \end{cases} \tag{54}$$

where

$$\begin{aligned} F &:= A + ZR^{-1}C_1 + (B_2 + ZR^{-1}D_1) \\ &\quad \times (K_1 - K_2 V^{-1/2} C_2) - (AQC_2' + B_1 D_2' \\ &\quad + ZR^{-1} C_1 Q C_2') V^{-1/2} C_2, \\ G &:= (B_2 + ZR^{-1}D_1) K_2 V^{-1/2} + (AQC_2' \\ &\quad + B_1 D_2' + ZR^{-1} C_1 Q C_2') V^{-1}, \\ H &:= K_1 - K_2 V^{-1/2} C_2, \\ J &:= K_2 V^{-1/2}, \end{aligned} \tag{55}$$

satisfies

$$\mathscr{C} \in \mathcal{A}_\infty(\mathcal{G}) \quad \text{and} \quad J(G, \hat{C}) = \mathrm{tr}\,(C_0 Q C_0') - \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0') + J(G_{fi}(Q), K).$$

To solve the output-feedback mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control problem using Theorem 5.2 the following steps should be followed.
(1) Check if $\mathcal{A}_\infty(\mathcal{G})$ is not empty. This can be done by applying the standard $\mathcal{H}_\infty$ theory for discrete-time systems and solving two discrete-time $\mathcal{H}_\infty$ Riccati equations as in Basar and Bernhard (1991), Iglesias and Glover (1991), Limebeer *et al.* (1989), Liu *et al.* (1991), and Stoorvogel (1990). If $\mathcal{A}_\infty(\mathcal{G})$ is not empty, let Q denote the unique solution to (47)–(50).
(2) Construct the auxiliary full-information plant $\mathcal{G}_{fi}(Q)$ defined in (51). Let $\epsilon > 0$ be given. Solve the convex program (27), corresponding to $\mathcal{G}_{fi}(Q)$, to compute a full-information gain $K = [K_1 \ K_2] \in \mathcal{A}_{\infty,m}(\mathcal{G}_{fi}(Q))$ such that $J(G_{fi}(Q), K) \le v(\mathcal{G}_{fi}(Q)) + \epsilon$. With this full-information gain, construct the output-feedback controller $\mathscr{C}$ in (54). Then $\mathscr{C}$ belongs to $\mathcal{A}_\infty(\mathcal{G})$, and satisfies $J(G, \hat{C}) \le v(\mathcal{G}) + \epsilon$.

Proof. It follows from Stoorvogel (1990), that the plant $\mathcal{G}$ given by (46) can be represented as the feedback interconnection shown in Fig. 5, where $Q \ge 0$ satisfies equations (47)–(50), and

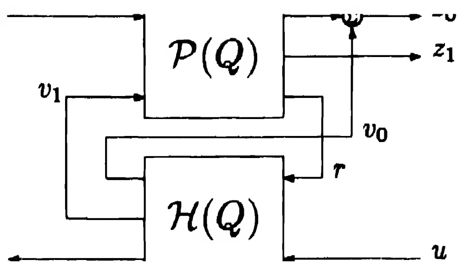


Fig 5 Equivalent representation of the transfer matrix G.

the plants $\mathcal{P}(Q)$ and $\mathcal{H}(Q)$ are defined by

$$\mathcal{P}(Q) := \begin{cases} \sigma x_p = (A - (AQC_2' + B_1 D_2') V^{-1} C_2) x_p \\ \quad + (B_1 - (AQC_2' + B_1 D_2') V^{-1} D_2) w \\ \quad - ZR^{-1/2} v_1 =: A_p x_p \\ \quad + B_{1p} w + B_{2p} v_1 \\ z_0 = C_0 x_p =: C_{0p} x_p \\ z_1 = (C_1 - C_1 Q C_2' V^{-1} C_2) x_p \\ \quad - C_1 Q C_2' V^{-1} D_2 w + R^{1/2} v_1 \\ \quad =: C_{1p} x_p + D_{1p} w + D_{2p} v_1 \\ r = V^{-1/2} C_2 x_p + V^{-1/2} D_2 w \\ \quad =: C_{2p} x_p + D_{21p} w, \end{cases} \tag{56}$$

and

$$\mathcal{H}(Q) := \begin{cases} \sigma x_g = A_g x_g + B_{1g} r + B_{2g} u \\ v_0 = C_{0g} x_g + D_{02g} u \\ v_1 = C_{1g} x_g + D_{11g} r + D_{12g} u \\ y = C_2 x_g + V^{1/2} r, \end{cases} \tag{57}$$

where the matrices in (57) are those introduced in the definition of $\mathcal{G}_{fi}(Q)$ given by (51). In (Stoorvogel, 1990), it has also been shown that $\mathscr{C} \in \mathcal{A}(\mathcal{G})$ if and only if $\mathscr{C}$ internally stabilizes the interconnection of $\mathcal{P}(Q)$ and $\mathcal{H}(Q)$. Routine algebra also shows that Q is the controllability gramian of the pair $(A_p, [B_{1p} \ B_{2p}])$, and that $\mathcal{P}(Q)$ satisfies the hypotheses of Lemma 5.1.

First, we show that

$$v(\mathcal{G}_{fi}(Q)) \le v(\mathcal{G}) - \mathrm{tr}\,(C_0 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0), \tag{58}$$

and that part 2 of the theorem holds. Suppose $\alpha > v(\mathcal{G})$. From the definition of $v(\mathcal{G})$ it follows that there exists a controller $\mathscr{C} \in \mathcal{A}_\infty(\mathcal{G})$ such that $J(G, C) < \alpha$. Apply the controller $\mathscr{C}$ to the interconnection of Fig. 5. Now using Lemma 5.1, since $\mathscr{C} \in \mathcal{A}_\infty(\mathcal{G})$ we get that $\mathscr{C} \in \mathcal{A}_\infty(\mathcal{H}(Q))$. Moreover,

$$J(G, C) = J(H(Q), C) + \mathrm{tr}\,(C_0 Q C_0') + 2\mathrm{tr}\,(D_0 D_c C_2 Q C_0'), \tag{59}$$

where $D_c = C(\infty)$ denotes the direct feedthrough term of the controller $\mathscr{C}$.

Define the full-information controller

$$C^* = C[C_2 \quad V^{1/2}],$$

and apply the controller $\mathscr{C}^*$ to the full-information plant $\mathscr{G}_{fi}(Q)$ defined in (51). Clearly, $\mathscr{C}^* \in \mathscr{A}_\infty(\mathscr{G}_{fi}(Q))$. Further, an easy calculation shows that

$$J(G_{fi}(Q), C^*) = J(H(Q), C) + 2\mathrm{tr}\,(D_0 D_c C_2 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0'). \quad (60)$$

From (59)–(60), we obtain

$$J(G_{fi}(Q), C^*) = J(G, C) - \mathrm{tr}\,(C_0 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0).$$

Now we can apply Theorem 4.1 to the full-information plant $\mathscr{G}_{fi}$ to conclude that there exists a static controller $\mathscr{K} \in \mathscr{A}_{\infty,m}(\mathscr{G}_{fi}(Q))$ such that

$$v(G_{fi}(Q)) \le J(G_{fi}(Q), K) \le J(G_{fi}(Q), C^*) < \alpha - \mathrm{tr}\,(C_0 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0).$$

This shows that part 2 of the theorem holds. Moreover, taking limit as $\alpha \to v(\mathscr{G})$, we may also conclude that (58) holds.

We now show that part 3 of the theorem holds, and that

$$v(G_{fi}(Q)) \ge v(G) - \mathrm{tr}\,(C_0 Q C_0') + \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0). \quad (61)$$

Here the key is to synthesize the given full-information control law $u = K_1 x_g + K_2 r$ using a dynamic feedback controller which uses only the output y for the system $\mathscr{H}(Q)$. In this construction, we essentially invert the transfer function from r to y in $\mathscr{H}(Q)$.

Define the dynamic controller

$$\hat{\mathscr{C}} := \begin{cases} \sigma x_c = A_g x_c + B_{2g} u + B_{1g} V^{-1/2}(y - C_2 x_c) \\ u = K_1 x_c + K_2 V^{-1/2}(y - C_2 x_c), \end{cases} \quad (62)$$

where K_1 and K_2 are such that $\mathscr{K} := [\mathscr{K}_1 \ \mathscr{K}_2] \in \mathscr{A}_\infty(\mathscr{G}_{fi}(Q))$, and A_g, B_g and V are defined in (51) and (47), respectively. Note that the controller (62) is an "observer based controller" for the auxiliary plant $\mathscr{H}(Q)$. Using this fact, and the stabilizing property of Q, it is easy to see that $\hat{\mathscr{C}} \in \mathscr{A}(\mathscr{H}(Q))$, and

$$T_{v_1 r}(H(Q), \hat{C}) = T_{v_1 r}(G_{fi}(Q), K).$$

Thus, $\hat{\mathscr{C}} \in \mathscr{A}_\infty(\mathscr{H}(Q))$.

Applying the controller $\hat{\mathscr{C}}$ to the interconnection of Fig. 5, it follows from Lemma 5.1 that $\hat{\mathscr{C}} \in \mathscr{A}_\infty(\mathscr{G})$. Furthermore, calculations as in first part of the proof imply that

$$v(\mathscr{G}) \le J(G, \hat{C}) = \mathrm{tr}\,(C_0 Q C_0') - \mathrm{tr}\,(C_0 Q C_2' V^{-1} C_2 Q C_0') + J(G_{fi}(Q), K).$$

This shows that part 3 holds. Moreover, by taking infimum over all full-information controllers $\mathscr{K} \in \mathscr{A}_\infty(\mathscr{G}_{fi}(Q))$, we may also conclude that (61) is satisfied. Now equations (58) and (61) together prove part 1 of the theorem. ■

6. CONCLUSIONS

In this paper we have considered a (sub-optimal) mixed $\mathscr{H}_2/\mathscr{H}_\infty$ control problem for discrete-time systems. This synthesis problem is well motivated since it represents a problem of (LQG) disturbance attenuation, as measured by the mixed $\mathscr{H}_2/\mathscr{H}_\infty$ performance measure, subject to a robust stability constraint.

We have shown that when the state of the plant, or the state and the exogenous input, is available for feedback, *memoryless feedback gains* offer the best possible performance. The optimal mixed $\mathscr{H}_2/\mathscr{H}_\infty$ performance, with state or full information feedback, was shown to be given by the value of a *finite-dimensional convex program*. This means that there are efficient numerical methods to compute the optimal performance, and a nearly optimal feedback gain. The reader may find an excellent description of some of these algorithms in Boyd and Barratt (1990). An ellipsoid algorithm has been developed in Rotea (1991) for a problem similar to the one considered in this paper.

In the case of output-feedback, it is shown that mixed $\mathscr{H}_2/\mathscr{H}_\infty$ controllers can be chosen to be a combination of an $\mathscr{H}_\infty$ *filter*, and a *full-information gain* for the mixed $\mathscr{H}_2/\mathscr{H}_\infty$ synthesis problem of a suitably constructed auxiliary plant. Thus, the output-feedback problem is no more difficult than the full-information problem. This appears to be the first complete solution to the mixed $\mathscr{H}_2/\mathscr{H}_\infty$ problem with dynamic output-feedback for discrete-time system.

Finally, the results in this paper may be combined with those of Rotea and Khargonekar (1991b) to solve mixed $\mathscr{H}_2/\mathscr{H}_\infty$ problems that involve time domain ℓ_∞ constraints.

REFERENCES

Bambang, R., E. Shimemura and K. Uchida (1990). Discrete-time mixed $\mathscr{H}_2/\mathscr{H}_\infty$ control and state estimation problems. Presented at *T90 Korean Automatic Control Conference*, Seoul, 26–27 October.

Basar, T. and P. Bernhard (1991). *$\mathscr{H}_\infty$-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Birkhauser, New York.

Bernussou, J., P. L. D. Peres and J. C. Jeromel (1989). A linear programming oriented procedure for quadratic stabilization of uncertain systems. *Systems and Control Letters*, **13**, 65–72.

Bernstein, D. S. and W. M. Haddad (1989). LQG control with an $\mathscr{H}_\infty$ performance bound: A Ricatti equation approach. *IEEE Trans. on Aut. Control*, **34**, 293–305.

Boyd, S. P. and C. Barratt (1990). *Linear Controller Design*. Prentice-Hall, Englewood Cliffs, NJ.

Dorato, P. (1991). A survey of multiobjective design techniques. In S. P. Battacharyya and L. H. Keel (Eds), *Control of Uncertain Dynamical Systems*, pp. 249–261. CRC Press, Boca Raton.

Doyle, J. C., K. Zhou and B. Bodenheimer (1989a). Optimal control with mixed $\mathcal{H}_2$ and $\mathcal{H}_\infty$ performance objectives. *Proc. of the American Control Conf.*, pp. 2065–2070.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989b). State-space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ and control problems. *IEEE Trans. on Aut. Control*, **34**, 831–847.

Haddad, W. M., D. S. Bernstein and D. Mustafa (1991). Mixed-norm $\mathcal{H}_2/\mathcal{H}_\infty$ regulation and estimation: The discrete-time case. *Systems and Control Letters*, **16**, 235–247.

Iglesias, P. A. and K. Glover (1991). State-space approach to discrete-time $\mathcal{H}_\infty$ control. *Int. J. Control*, **54**, 1031–1073.

Khargonekar, P. P. and M. A. Rotea (1991a). Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: a convex optimization approach. *IEEE Trans. on Aut. Control*, **36**, 824–837.

Khargonekar, P. P. and M. A. Rotea (1991b). Controller synthesis for multiple objective optimal control. In S. P. Battacharyya and L. H. Keel (Eds), *Control of Uncertain Dynamical Systems*. pp. 261–280. CRC Press, Boca Raton.

Limebeer, D. J. N., M. Green and D. Walker (1989). Discrete-time $\mathcal{H}_\infty$ control. *Proc. IEEE Conf. on Decision and Control*, pp. 392–396.

Liu, K. Z., T. Mita and H. Kimura (1991) Complete solution to the standard $\mathcal{H}_\infty$ control problem of discrete-time systems. Preprint.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York.

Molinari, B. P. (1975). The stabilizing solution of discrete algebraic Riccati equation. *IEEE Trans. on Aut. Control*, **20**, 396–399.

Mustafa, D. and D. S. Bernstein (1991). LQG cost bounds in discrete-time $\mathcal{H}_2/\mathcal{H}_\infty$ control. *Trans. of the Institute of Measurement and Control*, to appear.

Ran, A. C. M. and R. Vreugdenhil (1988). Existence and comparison theorems for algebraic Riccati equations for continuous and discrete time systems. *Linear Algebra and its Applications*, **99**, 63–83.

Rotea, M. A. (1990). Multiple objective and robust control of linear systems. Ph.D. dissertation, University of Minnesota.

Rotea, M. A. (1991). The generalized $\mathcal{H}_2$ control problem. *Proc. First IFAC Symposium on Design Methods of Control Systems*, Zurich, **1**, 112–117.

Rotea, M. A. and P. P. Khargonekar (1991a). $\mathcal{H}_1$-optimal control with an $\mathcal{H}_\infty$-constraint: the state-feedback case. *Automatica*, **27**, 307–316.

Rotea, M. A. and P. P. Khargonekar (1991b). Generalized $\mathcal{H}_2/\mathcal{H}_\infty$ control via convex optimization. *Proc. IEEE Conf. on Decision and Control*, pp. 2719–2720.

Steinbuch, M. and O. Bosgra (1991). Necessary conditions for static and fixed order dynamic mixed $\mathcal{H}_2/\mathcal{H}_\infty$ optimal control. *Proc. American Control Conf.*, pp. 1137–1142.

Stoorvogel, A. A. (1990). The $\mathcal{H}_\infty$ control problem: A state-space approach. Ph.D. Dissertation, University of Eindhoven.

Yeh, H., S. Banda and B. Chang (1992). Necessary and sufficient conditions for mixed $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control. *IEEE Trans. on Aut. Control*, **37**, 355–358.

Zhou, K., J. Doyle, K Glover and B. Bodenheimer (1990). Mixed $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control. *Proc. of the American Control Conf.*, pp. 2502–2507.

Automatica, Vol. 29, No. 1, pp. 71–109, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

The Complex Structured Singular Value*

A. PACKARD† and J. DOYLE‡

A tutorial introduction to the complex structured singular value (μ) is presented, with an emphasis on computable bounds and robust stability and performance tests for transfer functions and their state space realizations.

Key Words—Computational methods; control system analysis; disturbance rejection; frequency domain; matrix algebra; multivariable control systems; performance bounds; robust control; sensitivity analysis; state space methods.

Abstract—A tutorial introduction to the complex structured singular value (μ) is presented, with an emphasis on the mathematical aspects of μ. The μ-based methods discussed here have been useful for analysing the performance and robustness properties of linear feedback systems. Several tests for robust stability and performance with computable bounds for transfer functions and their state space realizations are compared, and a simple synthesis problem is studied. Uncertain systems are represented using Linear Fractional Transformations (LFTs) which naturally unify the frequency-domain and state space methods.

1. INTRODUCTION

THIS PAPER GIVES a fairly complete introduction to the Structured Singular Value (μ) for complex perturbations. This paper is intended to be of tutorial value on the mathematical aspects of μ, and it is assumed that the reader is familiar with the control engineering motivation. The μ-based methods discussed here have been useful for analysing the performance and robustness properties of linear feedback systems. The more elementary methods are now available in commercial software products and the manual (Balas *et al.*, 1991) for one such product would serve as a tutorial introduction to the engineering motivation. The interested reader might also consult the tutorial in Stein and Doyle (1991) or other application-oriented papers, such as Skogestad *et al.* (1988). We present very few new results in this paper, although many of the results have appeared only in reports and conference proceedings. The paper is reasonably self-contained, skipping only those proofs which are readily available in the literature.

Section 3 begins with the definition of μ and some of its elementary properties, including simple bounds that form the basis for computational schemes. This section also introduces the relationship between the upper bound for μ and Linear Matrix Inequalities (LMIs) which results in a simple characterization of the convexity properties of the upper bound. The connections between μ and Linear Fractional Transformations are introduced in Section 4. These connections, especially the Main Loop Theorem, form the basis for most of the applications of μ to linear systems. In Section 5, using the definition of μ, and the Main Loop theorem, robust stability and robust performance theorems are derived for linear systems with structured linear fractional uncertainty.

Section 6 covers a maximum-modulus theorem for linear fractional transformations. Section 7 presents a generalization of the standard power algorithms for computing the spectral radius or maximum singular value of a matrix to the computation of μ. This power algorithm provides an attractive method for computing lower bounds for μ. Sections 8 and 9 consider issues associated with the upper bound, focusing particularly on conditions under which the upper bound is equal to μ. For certain simple block structures, this equality is guaranteed.

The remainder of the paper discusses applications of μ to problems motivated by control systems. Section 10 considers how various μ problems can be viewed in transfer function and state-space formulations. This leads to a variety of tests for robust performance, each with an interesting and useful interpretation. In Section 11, many (computable) necessary and sufficient conditions for quadratic stability of uncertain systems are given, for a wide variety of uncertainty structures. The proof techniques used in each different case are identical, giving a unifying treatment of many known and new results.

* Received 13 February 1992; received in final form 23 July 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor H. Kimura.

† Mechanical Engineering, University of California, Berkeley, CA 94720, U.S.A.

‡ Electrical Engineering, 116-81, Caltech, Pasadena, CA 91125, U.S.A.

Section 12 considers a simple case of μ-synthesis, the problem of minimizing μ as a function of some free parameter, such as a controller. Not surprisingly, μ-synthesis is a much harder problem than μ-analysis. For example, unlike μ-analysis problems, no method for minimizing the upper bound for μ in the synthesis problem using convex optimization has been found.

Finally, the paper outlines some related work in Section 13, beginning with a brief history of the early development of the μ theory. This outline is not intended to be exhaustive or complete, but simply to touch on a few of the topics nearest to this paper that were not considered in detail. LMIs are presented as potentially unifying theoretical and computational tools. The relationship between μ and quadratic vs L_1 notions of robust performance and robust stability is then discussed, followed by μ with mixed real and complex perturbations. The section ends with a discussion of model validation and generalizations of μ.

2. NOTATION

The notation is standard. $\mathbf{R}$ denotes the set of real numbers; $\mathbf{C}$ denotes the set of complex numbers; $|\cdot|$ is the absolute value of elements in $\mathbf{R}$ or $\mathbf{C}$; $\mathbf{R}^n$ is the set of real n vectors; $\mathbf{C}^n$ is the set of complex n vectors; $\|v\|$ is the Euclidean norm for $v \in \mathbf{C}^n$, $\|v\|^2 := \sum_{i=1}^{n} |v_i|^2$; l_2^n denotes the set of square summable sequences in C^n; $\|e\|_2$ is the l_2 norm of sequence $e \in l_2^n$, $\|e\|_2^2 := \sum_{k=1}^{\infty} \|e_k\|^2$; $\mathbf{R}^{n\times m}$ is the set of $n \times m$ real matrices; $\mathbf{C}^{n\times m}$ is the set of $n \times m$ complex matrices; $\mathbf{H}^n$ is the set of Hermitian $n \times n$ complex matrices; I_n is a $n \times n$ identity matrix; and 0_n or $0_{n\times m}$ is an entirely zero matrix of obvious dimensions. For $M \in \mathbf{C}^{n\times m}$; M^T is the transpose of M; M^* is the complex conjugate transpose of M; $\underline{\sigma}(M)$ is the minimum singular value of M; $\sigma_i(M)$ is a singular value of M; $\bar{\sigma}(M)$ is the maximum singular value of M. For $M \in \mathbf{C}^{n\times n}$: $\lambda_i(M)$ is an eigenvalue of M; $\rho(M)$ is the spectral radius of M, $\rho(M) := \max_i |\lambda_i(M)|$; $\operatorname{tr}(M)$ is the trace of M, $\operatorname{tr}(M) := \sum_{i=1}^{n} M_{ii}$. If $M \in \mathbf{C}^{n\times n}$ satisfies $M = M^*$ then $M > 0$ denotes that M is positive definite, and $M^{1/2}$ denotes the unique positive definite Hermitian square root.

3. STRUCTURED SINGULAR VALUE

This section is devoted to defining the structured singular value, a matrix function denoted by $\mu(\cdot)$. We consider matrices $M \in \mathbf{C}^{n\times n}$. In the definition of $\mu(M)$, there is an underlying structure $\mathbf{\Delta}$, (a prescribed set of block diagonal matrices) on which everything in the sequel depends. This structure may be defined differently for each problem depending on the uncertainty and performance objectives of the problem. Defining the structure involves specifying three things: the total number of blocks, the type of each block, and their dimensions.

In this paper, we consider two types of blocks—repeated scalar and full blocks. Two nonnegative integers, S and F, denote the number of repeated scalar blocks and the number of full blocks, respectively. To bookkeep the block dimensions, we introduce positive integers $r_1, \ldots, r_S$; $m_1, \ldots, m_F$. The ith repeated scalar block is $r_i \times r_i$, while the jth full block is $m_j \times m_j$. With those integers given, define $\mathbf{\Delta} \subset \mathbf{C}^{n\times n}$ as

$$\mathbf{\Delta} = \{\operatorname{diag}[\delta_1 I_{r_1}, \ldots, \delta_S I_{r_S}, \Delta_{S+1}, \ldots, \Delta_{S+F}]:$$
$$\delta_i \in \mathbf{C}, \Delta_{S+j} \in \mathbf{C}^{m_j\times m_j}, 1 \le i \le S, 1 \le j \le F\}. \quad (3.1)$$

For consistency among all the dimensions, we must have $\sum_{i=1}^{S} r_i + \sum_{j=1}^{F} m_j = n$. Often, we will need norm bounded subsets of $\mathbf{\Delta}$, and we introduce the notation

$$\mathbf{B}_{\mathbf{\Delta}} = \{\Delta \in \mathbf{\Delta} : \bar{\sigma}(\Delta) \le 1\}. \quad (3.2)$$

Note that in (3.1) all of the repeated scalar blocks appear first and the full blocks are square. This is done to keep the notation as simple as possible and can easily be relaxed.

Definition 3.1. For $M \in \mathbf{C}^{n\times n}$, $\mu_{\mathbf{\Delta}}(M)$ is defined

$$\mu_{\mathbf{\Delta}}(M) := \frac{1}{\min\{\bar{\sigma}(\Delta) : \Delta \in \mathbf{\Delta}, \det(I - M\Delta) = 0\}}, \quad (3.3)$$

unless no $\Delta \in \mathbf{\Delta}$ makes $I - M\Delta$ singular, in which case $\mu_{\mathbf{\Delta}}(M) := 0$.

Remark 3.2. The set $\mathbf{\Delta}$ defines a multi-index of integers and vice versa, so it makes sense to identify as one object the set, the block structure, and the associated multi-index of integers and refer simply to a block structure $\mathbf{\Delta}$. Clearly, $\mu_{\mathbf{\Delta}}(M)$ depends on the block structure $\mathbf{\Delta}$ as well as the matrix M.

Remark 3.3. Without loss in generality, the full blocks in the minimal norm Δ can each be chosen to be dyads (rank = 1). To see this, first consider the case of only one full block, $\mathbf{\Delta} = \mathbf{C}^{n\times n}$. Suppose that $I - M\Delta$ is singular.

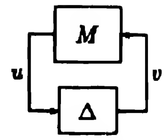


FIG. 1. M–Δ feedback connection.

Then for some unit-norm vector $x \in \mathbf{C}^n$, $M\Delta x = x$. Define $y := \Delta x$. It follows that $y \neq 0$, and $\|y\| \leq \bar{\sigma}(\Delta)$. Hence, define a new perturbation, $\tilde{\Delta} \in \mathbf{C}^{n \times n}$ as

$$\tilde{\Delta} := yx^*.$$

Note that $\bar{\sigma}(\tilde{\Delta}) = \|y\| \leq \bar{\sigma}(\Delta)$, and $y = \tilde{\Delta}x$, so that $I - M\tilde{\Delta}$ is also singular. Repeating this on a block-by-block basis allows for each full block to be a dyad.

Remark 3.4. It is instructive to consider a "feedback" interpretation of $\mu_\Delta(M)$ at this point. Let $M \in \mathbf{C}^{n \times n}$ be given, and consider the loop shown in Fig. 1. This picture is meant to represent the loop equations $u = Mv$, $v = \Delta u$. As long as $I - M\Delta$ is nonsingular, the only solutions u, v to the loop equations are $u = v = 0$. However, if $I - M\Delta$ is singular, then there are infinitely many solutions to the equations, and norms $\|u\|$, $\|v\|$ of the solutions can be arbitrarily large. Motivated by connections with stability of systems, which will be explored in detail in the sequel, we call this constant matrix feedback system "unstable". Likewise, the term "stable" will describe the situation when the only solutions are identically zero. In this context then, $\mu_\Delta(M)$ provides a measure of the smallest structured Δ that causes "instability" of the constant matrix feedback loop shown above. The norm of this "destabilizing" Δ is exactly $1/\mu_\Delta(M)$.

Remark 3.5. It is immediate from the definition that for any $\alpha \in \mathbf{C}$, $\mu(\alpha M) = |\alpha|\, \mu(M)$. However, for all nontrivial block structures, the function $\mu : \mathbf{C}^{n \times n} \to \mathbf{R}$ is not a norm, since it does not satisfy the triangle inequality.

Remark 3.6. A natural question is why we work with μ and not $1/\mu$, especially in view of equation (3.3) in the definition of μ. While it is clearly a matter of taste, there are important reasons. Mathematically, μ is continuous and bounded and scales as indicated above. Perhaps more importantly, it connects more naturally with LFTs and generalizes the spectral radius and maximum singular value, as will be seen below.

An alternative expression for $\mu_\Delta(M)$ follows easily from the definition.

Lemma 3.7. $\mu_\Delta(M) = \max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M)$.

Proof. Since for any $\alpha \in \mathbf{C}$, $\mu_\Delta(\alpha M) = |\alpha|\, \mu_\Delta(M)$, we need only consider two cases: $\mu_\Delta(M) = 1$ iff $\max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M) = 1$ and $\mu_\Delta(M) = 0$ iff $\max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M) = 0$. These facts can be verified directly from the definition.

This lemma applies continuity of the function $\mu : \mathbf{C}^{n \times n} \to \mathbf{R}$ based on continuity of the spectral radius and max functions, and the compactness of $\mathbf{B}_\Delta$.

We can relate $\mu_\Delta(M)$ to familiar linear algebra quantities when $\mathbf{\Delta}$ is one of two extreme sets:

- If $\mathbf{\Delta} = \{\delta I : \delta \in \mathbf{C}\}$ ($S = 1$, $F = 0$, $r_1 = n$), then $\mu_\Delta(M) = \rho(M)$, the spectral radius of M.

Proof. This follows immediately from Lemma 3.7.

- If $\mathbf{\Delta} = \mathbf{C}^{n \times n}$ ($S = 0$, $F = 1$, $m_1 = n$), then $\mu_\Delta(M) = \bar{\sigma}(M)$.

Proof. If $\bar{\sigma}(\Delta) < 1/\bar{\sigma}(M)$, then $\bar{\sigma}(M\Delta) < 1$, so $I - M\Delta$ is nonsingular. Applying equation (3.3) implies $\mu_\Delta(M) \leq \bar{\sigma}(M)$. On the other hand, let u and v be unit vectors satisfying $Mv = \bar{\sigma}(M)u$, and define $\Delta := (1/\bar{\sigma}(M))vu^*$. Then $\bar{\sigma}(\Delta) = 1/\bar{\sigma}(M)$ and $I - M\Delta$ is obviously singular. Hence, $\mu_\Delta(M) \geq \bar{\sigma}(M)$.

Obviously, for a general $\mathbf{\Delta}$ as in (3.1) we must have $\{\delta I_n : \delta \in \mathbf{C}\} \subset \mathbf{\Delta} \subset \mathbf{C}^{n \times n}$. Hence directly from the definition of μ, and the two special cases above, we conclude that

$$\rho(M) \leq \mu_\Delta(M) \leq \bar{\sigma}(M). \tag{3.4}$$

These bounds by themselves may prove little information on the value of μ, because the gap between ρ and $\bar{\sigma}$ can be arbitrarily large. They are refined with transformations on M that do not affect $\mu_\Delta(M)$, but do affect ρ and $\bar{\sigma}$. To do this, define two subsets of $\mathbf{C}^{n \times n}$

$$\mathbf{Q} = \{Q \in \mathbf{\Delta} : Q^*Q = I_n\}, \tag{3.5}$$

$$\mathbf{D} = \{\mathrm{diag}\,[D_1, \ldots, D_S, d_{S+1}I_{m_1}, \ldots, d_{S+F}I_{m_F}] : D_i \in \mathbf{C}^{r_i \times r_i}, D_i = D_i^* > 0, d_{S+j} \in \mathbf{R}, d_{S+j} > 0\}. \tag{3.6}$$

The reasons for taking $\mathbf{D}$ positive will be clear shortly. Note that for any $\Delta \in \mathbf{\Delta}$, $Q \in \mathbf{Q}$, and $D \in \mathbf{D}$,

$$\begin{gathered} Q^* \in \mathbf{Q}, \quad Q\Delta \in \mathbf{\Delta}, \quad \Delta Q \in \mathbf{\Delta}, \\ \bar{\sigma}(Q\Delta) = \bar{\sigma}(\Delta Q) = \bar{\sigma}(\Delta), \end{gathered} \tag{3.7}$$

$$D^{1/2}\Delta = \Delta D^{1/2}. \tag{3.8}$$

Theorem 3.8. For all $Q \in \mathbf{Q}$ and $D \in \mathbf{D}$

$$\mu_\Delta(MQ) = \mu_\Delta(QM) = \mu_\Delta(M) = \mu_\Delta(D^{1/2}MD^{-1/2}). \qquad (3.9)$$

Proof. For all $D \in \mathbf{D}$ and $\Delta \in \boldsymbol{\Delta}$,

$$\det(I - M\Delta) = \det(I - MD^{-1/2}\Delta D^{1/2}) = \det(I - D^{1/2}MD^{-1/2}\Delta),$$

since D commutes with Δ. Therefore $\mu_\Delta(D^{1/2}MD^{-1/2})$. Also, for each $Q \in \mathbf{Q}$, $\det(I - M\Delta) = 0$ if and only if $\det(I - MQQ^*\Delta) = 0$. Since $Q^*\boldsymbol{\Delta} = \boldsymbol{\Delta}$ and $\bar\sigma(Q^*\Delta) = \bar\sigma(\Delta)$, we get $\mu_\Delta(MQ) = \mu_\Delta(M)$ as desired. The argument for QM is the same.

Therefore, the bounds in (3.4) can be tightened to

$$\max_{Q \in \mathbf{Q}} \rho(QM) \le \max_{\Delta \in \mathbf{B}\boldsymbol{\Delta}} \rho(\Delta M) = \mu_\Delta(M) \le \inf_{D \in \mathbf{D}} \bar\sigma(D^{1/2}MD^{-1/2}), \qquad (3.10)$$

where the equality comes from Lemma 3.7. Note that in computing the infimum, and one of the diagonal entries of the elements of $\mathbf{D}$ can be assumed to be equal to 1. This is without loss in generality, since for any nonzero scalar γ, $D^{1/2}MD^{-1/2} = (\gamma D)^{1/2}M(\gamma D)^{-1/2}$. Hence, from this point on, we assume that $d_{S+F} \equiv 1$. Also, using the polar decomposition theorem for invertible matrices, it is easy to see that restricting the elements of $\mathbf{D}$ to be Hermitian, positive definite, as opposed to just invertible, does not affect the infimum. Certain convexity properties make the upper bound computationally attractive. For block structures with $S = 0$, it is shown by Safonov and Doyle (1984) that by using an exponential parametrization of $\mathbf{D}$, the function $\bar\sigma(D^{1/2}MD^{-1/2})$ is convex in $\log(D^{1/2})$. In Sezginer and Overton (1990) a very elegant and simple proof shows that the function $\bar\sigma(e^XMe^{-X})$ is convex on any convex set of commuting matrices $\mathbf{X}$. This generalizes the results in Safonov and Doyle (1984) and relies only on elementary linear algebra. The simplest convexity property is given in the following theorem, which shows that the function $\bar\sigma(D^{1/2}MD^{-1/2})$ has convex level sets.

Theorem 3.9. *Let* $M \in \mathbf{C}^{n \times n}$ be given, along with a scaling set $\mathbf{D}$, and $\beta > 0$. Then

$$\{D \in \mathbf{D} : (D^{1/2}MD^{-1/2}) < \beta\},$$

is convex.

Proof. The following chain of equivalences comprises the proof:

$$\begin{aligned}
&\bar\sigma(D^{1/2}MD^{-1/2}) < \beta \\
&\Leftrightarrow \lambda_{\max}(D^{-1/2}M^*D^{1/2}D^{1/2}MD^{-1/2}) < \beta^2 \\
&\Leftrightarrow D^{-1/2}M^*D^{1/2}D^{1/2}MD^{-1/2} - \beta^2 I < 0 \\
&\Leftrightarrow M^*DM - \beta^2 D < 0. \qquad (3.11)
\end{aligned}$$

The latter is clearly a convex condition in D.

Remark 3.10. The final condition in equation (3.11) is called a Linear Matrix Inequality (LMI). Note that although it is equivalent to the condition in Theorem 3.9, the functional dependence on D is much simpler and makes the convexity property clearer. For these reasons, LMIs appear to be attractive for computation. General linear matrix inequalities are discussed in greater detail in Section 13.2.

4 LINEAR FRACTIONAL TRANSFORMATIONS AND μ

The use of μ in control theory depends to a great extent on its intimate relationship with a class of general linear feedback loops called Linear Fractional Transformations (LFTs). This section explores this relationship with some simple theorems that can be obtained almost immediately from the definition of μ. To introduce these, consider a complex matrix M partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \qquad (4.1)$$

and suppose there is a defined block structure $\boldsymbol{\Delta}_2$ which is compatible in size with M_{22} (for any $\Delta_2 \in \boldsymbol{\Delta}_2$, $M_{22}\Delta_2$ is square). For $\Delta_2 \in \boldsymbol{\Delta}_2$, consider the loop equations

$$\begin{aligned}
e &= M_{11}d + M_{12}w, \\
z &= M_{21}d + M_{22}w, \qquad (4.2) \\
w &= \Delta_2 z,
\end{aligned}$$

which correspond to the block diagram in Fig. 2.

Definition 4.1. Given complex matrices M and Δ_2 as described above. This set of equations (4.2) is called well posed if for any vector d, there exist unique vectors w, z, and e satisfying the loop equations.

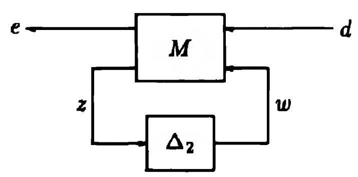


FIG. 2. Linear fractional transformation

It is easy to see that the set of equations is well posed if and only if the inverse of $I - M_{22}\Delta_2$ exists. If this inverse does not exist, then depending on d and M, there is either no solution to the loop equations, or there are an infinite number of solutions. When the inverse does exist, the vectors e and d satisfy $e = \mathcal{S}(M, \Delta_2)d$, where

$$\mathcal{S}(M, \Delta_2) := M_{11} + M_{12}\Delta_2(I - M_{22}\Delta_2)^{-1}M_{21}, \tag{4.3}$$

$\mathcal{S}(M, \Delta_2)$ is called a Linear Fractional Transformation (LFT). If $\mathbf{\Delta}_1$ is a block structure compatible in dimension with M_{11}, then for $\Delta_1 \in \mathbf{\Delta}_1$ an analogous formula describes $\mathcal{S}(\Delta_1, M)$,

$$\mathcal{S}(\Delta_1, M) := M_{22} + M_{21}\Delta_1(I - M_{11}\Delta_1)^{-1}M_{12},$$

where the upper loop of M is closed with a matrix Δ_1.

The $\mathcal{S}(M, \Delta_2)$ and $\mathcal{S}(\Delta_1, M)$ notation can be somewhat confusing on first encounter. It comes from the "star-product" of Redheffer. Suppose that Q and M are complex matrices, partitioned as

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}, \quad M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

with the matrix product $Q_{22}M_{11}$ well defined and square. If $I - Q_{22}M_{11}$ is invertible, then the block diagram in Fig. 3 is well defined. We can extend the definition of $\mathcal{S}$ so that it equals the result of this interconnection,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathcal{S}(Q, M)\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Simple manipulation gives

$$\mathcal{S}(Q, M) := \begin{bmatrix} \mathcal{S}(Q, M_{11}) & Q_{12}(I - M_{11}Q_{22})^{-1}M_{12} \\ M_{21}(I - Q_{22}M_{11})^{-1}Q_{21} & \mathcal{S}(Q_{22}, M) \end{bmatrix},$$

where $\mathcal{S}(Q, M_{11})$ and $\mathcal{S}(Q_{22}, M)$ are defined as above. Note that this definition is dependent on the partitioning of the matrices Q and M above; it may be well defined for one partition and not well defined for another.

For clarity, the notation $\mathcal{S}(\cdot, \cdot)$ should have two additional arguments, specifying dimensions of the partitions. However, in this paper, the only situations using $\mathcal{S}$ that will arise are LFTs, namely:

(1) the number of rows of Q is less than the number of columns of M, and the number of columns of Q is smaller than the number of rows of M, or:
(2) the number of columns of M is less than the number of rows of Q, and the number of rows of M is smaller than the number of columns of Q,

and all inputs/outputs into the (dimensionally) smaller matrix are closed in the interconnecting transformation. Hence, we do not need to specify the dimensions of the interconnecting channels, since they are equal to the dimension of the smaller matrix.

Alternative notation for LFTs has been used in previous papers, most notably

$$\mathcal{S}(M, \Delta_2) = \mathcal{F}_l(M, \Delta_2),$$
$$\mathcal{S}(\Delta_1, M) = \mathcal{F}_u(M, \Delta_1),$$

where l and u indicate that the lower and upper loop, respectively, are closed. We believe that the $\mathcal{S}$ notation is more natural, easier to work with, and generalizes smoothly to $\mathcal{S}(M, Q)$ in Fig. 3.

4.1. *Examples of LFTs*

Given the state space realization of a discrete time system

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} x_k \\ u_k \end{bmatrix} = M\begin{bmatrix} x_k \\ u_k \end{bmatrix}, \tag{4.4}$$

then its transfer matrix is

$$G(z) = D + C(zI - A)^{-1}B = \mathcal{S}\left(\frac{1}{z}I, M\right).$$

Systems with uncertainty can also be easily represented using LFTs. One natural type of uncertainty is unknown coefficients in a state space model. As a simple example, we will begin with a familiar idealized mass–spring–damper systems shown in Fig. 4. Suppose m, c, and k are fixed but uncertain, with $m = \bar{m}(1 + w_m\delta_m)$, $c = \bar{c}(1 + w_c\delta_c)$, $k = \bar{k}(1 + w_k\delta_k)$. Then defining $x_1 = y$ and $x_2 = \dot{y}$ we can write the differential

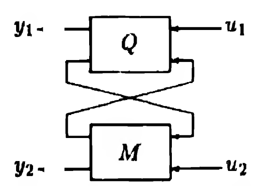


FIG. 3. General star product.

$$\ddot{y} + \frac{c}{m}\dot{y} + \frac{k}{m}y = \frac{F}{m}$$

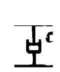


FIG. 4 Mass–spring–damper system.

equation in state space form as

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \mathcal{S}(M, \Delta)\begin{bmatrix} x \\ F \end{bmatrix}, \quad \Delta = \operatorname{diag}(\delta_m, \delta_c, \delta_k),$$

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{-\bar{k}}{\bar{m}} & \frac{-\bar{c}}{\bar{m}} & \frac{1}{\bar{m}} & -w_m & \frac{-w_c}{\bar{m}} & \frac{-w_k}{\bar{m}} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{-\bar{k}}{\bar{m}} & \frac{-\bar{c}}{\bar{m}} & \frac{1}{\bar{m}} & -w_m & \frac{-w_c}{\bar{m}} & \frac{-w_k}{\bar{m}} \\ 0 & \bar{c} & 0 & 0 & 0 & 0 \\ \bar{k} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

More generally, the perturbed state space system

$$x_{k+1} = A(\delta)x_k + B(\delta)d_k,$$
$$e_k = C(\delta)x_k + D(\delta)d_k,$$

where δ is a vector of parameters that enter rationally can be written as an LFT on a diagonal matrix Δ made up of the elements of δ, possibly repeated. The form of the LFT is Morton and McAfoos (1985)

$$\begin{bmatrix} x_{k+1} \\ e_k \\ z_k \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}\begin{bmatrix} x_k \\ d_k \\ w_k \end{bmatrix}$$

with perturbation $w_k = \Delta z_k$ yielding

$$\begin{bmatrix} x_{k+1} \\ e_k \end{bmatrix} = \mathcal{S}(M, \Delta)\begin{bmatrix} x_k \\ d_k \end{bmatrix}.$$

In general, for problems of this type it is easy to obtain realizations, but it is difficult to insure that they are minimal, except in the case where the parameters enter linearly.

A fundamental property of LFTs that contributes to their importance in linear systems theory is that interconnections of LFTs are again LFTs. For example, consider a situation with three components, each with a LFT uncertainty model. The interconnection is shown in Fig. 5. By simply reorganizing the diagram, collecting all of the known systems together, and collecting all of the perturbations (the Δ_is) together, we end up with the diagram in Fig. 6, where P

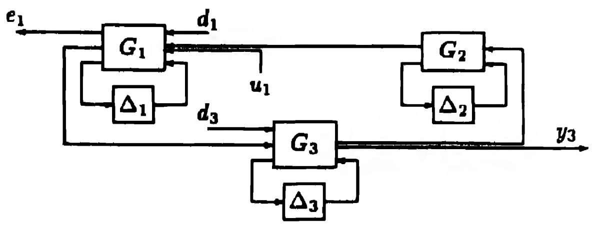


FIG. 5. Example interconnection of LFTs

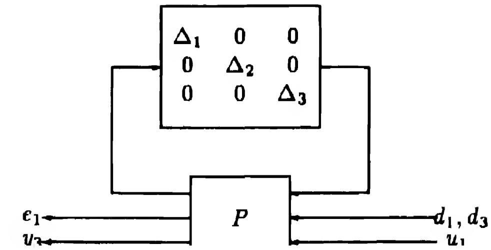


FIG. 6. Macroscopic representation of Fig. 5.

depends only on G_1, G_2, G_3 and the diagram layout. Note how general uncertainty at the component level becomes structured uncertainty at the system level. Additional information on LFTs and how they arise in engineering problems is found in Doyle *et al.* (1991).

4.2. *The Main Loop Theorem*

For notational ease, let $\mathbf{B}_i := \{\Delta_i \in \mathbf{\Delta}_i : \bar{\sigma}(\Delta_i) \le 1\}$. In this formulation, the matrix $M_{11} = \mathcal{S}(M, 0)$ may be thought of as the nominal map and $\Delta_2 \in \mathbf{B}_2$ viewed as a norm bounded perturbation from an allowable perturbation class, $\mathbf{\Delta}_2$. The matrices M_{12}, M_{21}, and M_{22} and the formula $\mathcal{S}(M, \cdot)$ reflect prior knowledge on how the unknown perturbation affects the nominal map, M_{11}. This type of uncertainty, called linear fractional, is natural for many control problems, and encompasses many other special cases considered by researchers in robust control and matrix perturbation theory.

The constant matrix problem to solve is: determine whether the LFT is well posed for all Δ_2 in $\mathbf{B}_2$ and, if so, then determine how "large" $\mathcal{S}(M, \Delta_2)$ can get for $\Delta_2 \in \mathbf{B}_2$.

Define a third structure $\mathbf{\Delta}$ as

$$\mathbf{\Delta} := \left\{\begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} : \Delta_1 \in \mathbf{\Delta}_1, \Delta_2 \in \mathbf{\Delta}_2\right\}. \tag{4.5}$$

Now there are three structures with respect to which we may compute μ. The notation we use to keep track is as follows: $\mu_1(\cdot)$ is with respect to $\mathbf{\Delta}_1$, $\mu_2(\cdot)$ is with respect to $\mathbf{\Delta}_2$, $\mu_{\mathbf{\Delta}}(\cdot)$ is with respect to $\mathbf{\Delta}$. In view of this, $\mu_1(M_{11})$, $\mu_2(M_{22})$ and $\mu_{\mathbf{\Delta}}(M)$ are all defined, though for instance, $\mu_1(M)$ is not defined. The first theorem follows immediately from the definition of μ.

Theorem 4.2. The linear fractional transformation $\mathcal{S}(M, \Delta_2)$ is well posed for all $\Delta_2 \in \mathbf{B}_2$ if and only if $\mu_2(M_{22}) < 1$.

As the perturbation Δ_2 deviates from zero, the matrix $\mathcal{S}(M, \Delta_2)$ deviates from M_{11}. The range of values that $\mu_1(\mathcal{S}(M, \Delta_2))$ takes on is

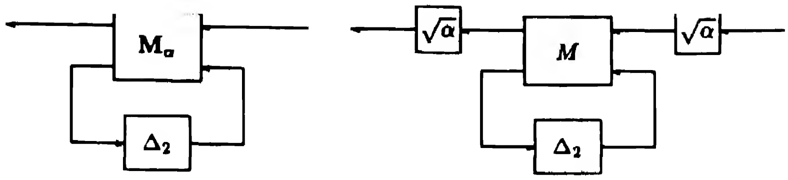


FIG. 7. Scaling for main loop theorem

intimately related to $\mu_{\Delta}(M)$, as follows:

Theorem 4.3. (Main Loop theorem)

$$\mu_{\Delta}(M) < 1 \Leftrightarrow \begin{array}{l} \mu_2(M_{22}) < 1 \\ \max\limits_{\Delta_2 \in \mathbf{B}_2} \mu_1(\mathcal{S}(M, \Delta_2)) < 1. \end{array}$$

Proof. First note that $\mu_{\Delta}(M) < 1$ implies that $\mu_2(M_{22}) < 1$, so we may assume the latter and prove the equivalence of the two remaining conditions. Let $\Delta_1 \in \mathbf{\Delta}_1$ be given, with $\bar{\sigma}(\Delta_1) \leq 1$, and define $\Delta = \operatorname{diag}[\Delta_1, \Delta_2]$ so that $\Delta \in \mathbf{\Delta}$. Now

$$\det(I - M\Delta) = \det \begin{bmatrix} I - M_{11}\Delta_1 & -M_{12}\Delta_2 \\ -M_{21}\Delta_1 & I - M_{22}\Delta_2 \end{bmatrix}. \tag{4.6}$$

By hypothesis $I - M_{22}\Delta_2$ is invertible, hence

$$\det(I - M\Delta) = \det(I - M_{22}\Delta_2) \cdot \det(I - M_{11}\Delta_1 - M_{12}\Delta_2(I - M_{22}\Delta_2)^{-1}M_{21}\Delta_1).$$

Collecting the Δ_1 terms leaves

$$\det(I - M\Delta) = \det(I - M_{22}\Delta_2)\det(I - \mathcal{S}(M, \Delta_2)\Delta_1). \tag{4.7}$$

By the definition of μ, the left-hand side of (4.7) is nonzero for all $\Delta \in \mathbf{B}_{\Delta}$ iff $\mu_{\Delta}(M) < 1$. Similarly, the right hand side is nonzero for all $\Delta = \operatorname{diag}[\Delta_1, \Delta_2] \in \mathbf{B}_{\Delta}$ iff $\mu_1(\mathcal{S}(M, \Delta_2)) < 1$ for all $\Delta_2 \in \mathbf{B}_2$. This completes the proof.

Remark 4.4. This theorem forms the basis for most uses of μ in linear system robustness analysis, whether from a state space, frequency domain, or Lyapunov approach.

Remark 4.5. This theorem is stated in terms of feasibility conditions, testing whether some quantity is less than one. This allows for an elegant statement and proof, but other versions are possible, with some complication in notation. Scaled versions of the Main Loop Theorem appear later in this section.

Remark 4.6. The importance of the theorem can be highlighted by a slight restatement. Suppose a property $\mathcal{P}$, of a matrix W can be related to a μ test on the matrix. That is, there exists some block structure $\mathbf{\Delta}_{\mathcal{P}}$ such that

$$\text{matrix } W \text{ satisfies property } \mathcal{P} \Leftrightarrow \mu_{\mathbf{\Delta}_{\mathcal{P}}}(W) < 1.$$

Then the perturbed matrix $\mathcal{S}(M, \Delta)$ is well defined, and has the property $\mathcal{P}$ for every $\Delta \in \mathbf{B}_{\Delta}$ if and only if $\mu_{\tilde{\Delta}}(M) < 1$, where $\tilde{\mathbf{\Delta}} := \{\operatorname{diag}[\Delta_{\mathcal{P}}, \Delta] : \Delta_{\mathcal{P}} \in \mathbf{\Delta}_{\mathcal{P}}, \Delta \in \mathbf{\Delta}\}$. In other words, whenever a property of a matrix can be related to a μ test, then there will be a μ test of greater complexity to determine if the property is robust to structured perturbations in the form of LFTs.

The role of the block structure $\mathbf{\Delta}_2$ is clear in this theorem—it is the structure for the original perturbation. However the role of the perturbation structure $\mathbf{\Delta}_1$ is often misunderstood. Note that $\mu_1(\cdot)$ appears on the right hand side of the theorem, so that the set $\mathbf{\Delta}_1$ defines what function of the matrix $\mathcal{S}(M, \Delta_2)$ is to be computed.

This theorem can be illustrated by a system-theoretic example with a transfer function and its state space realization. This example involves two of the simplest cases of LFTs. Suppose that $\mathbf{\Delta}_1 := \{\delta_1 I_n : \delta_1 \in \mathbf{C}\}$ and $\mathbf{\Delta}_2 = \mathbf{C}^{m \times m}$, which are the special cases considered in Section 3. Recall that for $A \in \mathbf{C}^{n \times n}$, $\mu_1(A) = \rho(A)$, and for $D \in \mathbf{C}^{m \times m}$, $\mu_2(D) = \bar{\sigma}(D)$. Now, let $\mathbf{\Delta}$ be the diagonal augmentation $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$, namely namely

$$\mathbf{\Delta} := \left\{ \begin{bmatrix} \delta_1 I_n & 0_{n \times m} \\ 0_{m \times n} & \Delta_2 \end{bmatrix} : \delta_1 \in \mathbf{C}, \Delta_2 \in \mathbf{C}^{m \times m} \right\} \subset \mathbf{C}^{(n+m) \times (n+m)}$$

Let $A \in \mathbf{C}^{n \times n}$, $B \in \mathbf{C}^{n \times m}$, $C \in \mathbf{C}^{m \times n}$, and $D \in \mathbf{C}^{m \times m}$, be given, and interpret them as the state space model of a discrete time system

$$x_{k+1} = Ax_k + Bu_k,$$
$$y_k = Cx_k + Du_k,$$

and let $M \in \mathbf{C}^{(n+m) \times (n+m)}$ be block state state matrix of the system,

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Applying the theorem with these data implies that the following are equivalent.

- The spectral radius of A satisfies $\rho(A) < 1$, and
$$\max_{\substack{\delta_1 \in \mathbf{C} \\ |\delta_1| \le 1}} (D + C\delta_1(I - A\delta_1)^{-1}B) < 1.$$
- The maximum singular value of D satisfies $\bar{\sigma}(D) < 1$, and
$$\max_{\substack{\Delta_2 \in \mathbf{C}^{m \times m} \\ \bar{\sigma}(\Delta_2) \le 1}} \rho(A + B\Delta_2(I - D\Delta_2)^{-1}C) < 1.$$
- The structures singular value of M satisfies $\mu_{\mathbf{\Delta}}(M) < 1$.

The first condition implies two things: the system is stable, and the $\|\cdot\|_\infty$ norm on the transfer function from u to y (obtained by setting $\delta_1 = 1/z$) is less than one. That is

$$\|G\|_\infty := \max_{\substack{z \in \mathbf{C} \\ |z| \ge 1}} \bar{\sigma}(D + C(zI - A)^{-1}B)$$
$$= \max_{\substack{\delta_1 \in \mathbf{C} \\ |\delta| \le 1}} \bar{\sigma}(D + C\delta_1(I - A\delta_1)^{-1}B) < 1.$$

The second condition implies that $(I - D\Delta_2)^{-1}$ is well defined for all $\bar{\sigma}(\Delta_2) \le 1$, and that the uncertain difference equation

$$x_{k+1} = (A + B\Delta_2(I - D\Delta_2)^{-1}C)x_k,$$

is stable for all such Δ_2.

The equivalence between the small gain condition, $\|G\|_\infty < 1$, and the stability robustness of the uncertain difference equation is well known as the small gain theorem, in its necessary and sufficient form for linear time invariant systems. What is important to see is that both of these conditions are in fact equivalent to one condition on the structured singular value. Already we have seen that the spectral radius and maximum singular value are special cases of μ. Here we see that additional important linear system properties, namely robust stability and input–output gain are also related to a particular case of the structured singular value.

Returning to the main loop theorem, note that the bound on the performance is the same as the bound on the perturbation, namely one. Scaling the matrix M by $1/\beta$, for some positive scalar β, and then applying the theorem gives the following.

Corollary 4.7. Let $\beta > 0$ be given. Then

$$\mu_{\mathbf{\Delta}}(M) < \beta \Leftrightarrow \begin{cases} \mu_2(M_{22}) < \beta \\ \max_{\Delta_2 \in (1/\beta)\mathbf{B}_2} \mu_1(\mathcal{S}(M, \Delta_2)) < \beta. \end{cases}$$

The bound on performance and the bound on the perturbation are related, they are reciprocals. For nonreciprocal values, certain blocks of M must be scaled and μ recomputed. Specifically, for $\alpha \ge 0$, define $\mathbf{M}_\alpha$ as

$$\mathbf{M}_\alpha = \begin{bmatrix} \alpha M_{11} & \sqrt{\alpha}\, M_{12} \\ \sqrt{\alpha}\, M_{21} & M_{22} \end{bmatrix}. \tag{4.8}$$

Some simple facts about $\mathbf{M}_\alpha$:

(1) if $\alpha = 0$ then $\mu_{\mathbf{\Delta}}(\mathbf{M}_\alpha) = \mu_2(M_{22})$;
(2) for any $\Delta_2 \in \mathbf{\Delta}_2$, with $I - M_{22}\Delta_2$ nonsingular, $\mathcal{S}(\mathbf{M}_\alpha, \Delta_2) = \alpha\mathcal{S}(M, \Delta_2)$;
(3) $\max\{\alpha\mu_1(M_{11}), \mu_2(M_{22})\} \le \mu_{\mathbf{\Delta}}(\mathbf{M}_\alpha) \le \max\{1, \alpha\}\mu_{\mathbf{\Delta}}(M)$;
(4) $\mu_{\mathbf{\Delta}}(\mathbf{M}_\alpha)$ is a continuous, nondecreasing function of α.

Theorem 4.8. Let $\beta > \mu_2(M_{22})$ be given, and $\alpha_\beta := \max\{\alpha > 0 : \mu_{\mathbf{\Delta}}(\mathbf{M}_\alpha) = \beta\}$. Then

$$\max_{\Delta_2 \in (1/\beta)\mathbf{B}_2} \mu_1(\mathcal{S}(M, \Delta_2)) = \frac{\beta}{\alpha_\beta}. \tag{4.9}$$

4.3. *Upper bound LFT results*

Theorem 4.3 gives necessary and sufficient conditions for performance/robustness characteristics in terms of a μ evaluation. The μ test always takes on the form "is $\mu(M) < 1$?". Hence, upper and lower bounds on μ can be used in the following manner: an upper bound gives a sufficient condition for the robustness/performance characteristic of the theorem; a lower bound gives a sufficient condition for when the robustness/performance will not be met. Clearly, both are important. The upper bound guarantees robustness of a property of a linear fractional transformation for perturbations up to a certain size, and a lower bound exhibits perturbations which cause a degree of degradation in the LFT's properties.

The above comments apply for any upper and lower bound. Of specific interest is the additional information that is obtained in using the $\bar{\sigma}(D^{1/2}MD^{-1/2})$ upper bound. Generally $\bar{\sigma}(D^{1/2}MD^{-1/2}) < 1$ implies a great deal more than $\mu_{\mathbf{\Delta}}(M) < 1$. As usual, let $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ be two given structures, and let $\mathbf{\Delta} = \{\mathrm{diag}\,[\Delta_1, \Delta_2] : \Delta_i \in \mathbf{\Delta}_i\}$. Similarly, let $\mathbf{D}_i$ be the appropriate D scaling sets for the two structures, (equation (3.6)) and denote $\mathbf{D}$ as the diagonal augmentation of these two sets, $\mathbf{D} := \{\mathrm{diag}(D_1, D_2) : D_i \in \mathbf{D}_i\}$.

Lemma 4.9 (Redheffer, 1959, 1960). Let M be given as in equation (4.1). Suppose there is a $D \in \mathbf{D}$ such that $\bar{\sigma}(D^{1/2}MD^{-1/2}) < 1$. Then there exists a $D_1 \in \mathbf{D}_1$ such that

$$\max_{\Delta_2 \in \mathbf{B}_2} \bar{\sigma}(D_1^{1/2}\mathcal{S}(M, \Delta_2)D_1^{-1/2}) < 1.$$

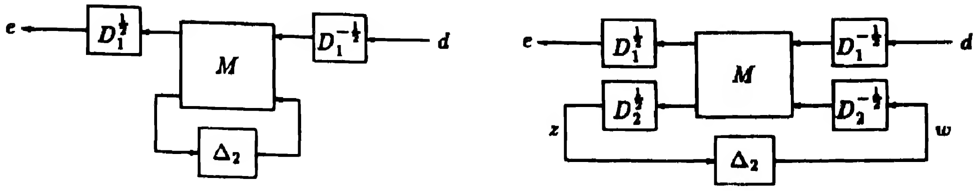


FIG. 8. Equivalent LFTs.

Proof. The easiest method of proof is just to track the norms of the various vectors in the loop equations for the linear fractional transformations shown in Fig. 8. Let D_1 and D_2 be the separate parts of the $D \in \mathbf{D}$ which achieves $\bar{\sigma}(D^{1/2}MD^{-1/2}) < 1$. Obviously, $\mu_2(M_{22}) < 1$, so for any $\Delta_2 \in \mathbf{B}_2$ the two LFTs are well posed, and from d to e are the same. Let $d \neq 0$ be any given complex vector of appropriate dimension, and let e, w, and z be the unique solutions to the loop equations for the linear fractional transformation on the right in Fig. 8. By hypothesis, we have

$$\|z\|^2 + \|e\|^2 < \|w\|^2 + \|d\|^2 \qquad (4.10)$$

and since $\bar{\sigma}(\Delta_2) \leq 1$

$$\|w\|^2 \leq \|z\|^2. \qquad (4.11)$$

Combining these gives

$$\|e\|^2 < \|d\|^2. \qquad (4.12)$$

Equation (4.12) also holds for the linear fractional transformation on the left, since the matrix relating d to e is the same for both linear fractional transformations. This implies that $\bar{\sigma}(D_1^{1/2}\mathcal{S}(M, \Delta_2)D_1^{-1/2}) < 1$ as desired.

Consider the problem determining the value of

$$\inf_{D_1 \in \mathbf{D}_1} \max_{\Delta_2 \in \mathbf{B}_2} \bar{\sigma}(D_1^{1/2}\mathcal{S}(M, \Delta_2)D^{-1/2}), \qquad (4.13)$$

and also finding a $D_1 \in \mathbf{D}_1$ that achieves a cost arbitrarily close to the infimum. Suppose the dimension of M_{11} is $n \times n$. Define an additional structure

$$\mathbf{\Delta}_N := \{\mathrm{diag}\,[\Delta, \Delta_2] : \Delta \in \mathbf{C}^{n \times n}, \Delta_2 \in \mathbf{\Delta}_2\}. \qquad (4.14)$$

Theorem 4.10. Let M, $\mathbf{\Delta}_2$, $\mathbf{D}_1$, and $\mathbf{\Delta}_N$ be given as above. Suppose that $\mu_{22}(M_{22}) < 1$. Define $\bar{\alpha}$ by

$$\bar{\alpha} = \sup_{\alpha > 0} \left\{ \alpha : \inf_{D_1 \in \mathbf{D}_1} \mu_{\mathbf{\Delta}_N}\left(\begin{bmatrix} D_1^{1/2} & 0 \\ 0 & I \end{bmatrix} \times \begin{bmatrix} \alpha M_{11} & \sqrt{\alpha} M_{12} \\ \sqrt{\alpha} M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} D_1^{-1/2} & 0 \\ 0 & I \end{bmatrix} \right) < 1 \right\}. \qquad (4.15)$$

Then

$$\inf_{D_1 \in \mathbf{D}_1} \max_{\Delta_2 \in \mathbf{B}_2} \bar{\sigma}(D_1^{1/2}\mathcal{S}(M, \Delta_2)D_1^{-1/2}) = \frac{1}{\bar{\alpha}}. \qquad (4.16)$$

In this section, all of the results were stated for $\mathcal{S}(M, \Delta_2)$. Analogous results hold for $\mathcal{S}(\Delta_1, M)$.

5. ROBUSTNESS TESTS WITH μ

The structured singular value can be used to quantify robustness margins for a linear system with linear fractional uncertainty. Specifically, suppose that $P(s)$ is a rational, proper matrix of size $(n_1 + n_2) \times (n_1 + n_2)$ and block structures $\mathbf{\Delta}_1 \subset \mathbf{C}^{n_1 \times n_1}$ and $\mathbf{\Delta}_2 \subset \mathbf{C}^{n_2 \times n_2}$ are given. Partition $P(s)$ as

$$P(s) = \begin{bmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{bmatrix}.$$

For $\Delta_1 \in \mathbf{\Delta}_1$, consider the interconnection shown in Fig. 9. For any $\Delta_1 \in \mathbf{B}_1$, $\mathcal{S}(\Delta_1, P(s))$ is the transfer function from $d_3 \rightarrow e_3$. The closed-loop system is said to be:

- well-posed if $\det(I - P_{11}(\infty)\Delta_1) \neq 0$. This is the necessary and sufficient condition that all closed-loop transfer in Fig. 5.1 be proper;
- stable if all closed-loop transfer functions in Fig. 5.1 are analytic in the closed right-half-plane.

Theorem 5.1. Suppose that $P(s)$ has all of its poles in the open left-half plane. Let $\beta > 0$. Then:

(1) For all $\Delta_1 \in \mathbf{\Delta}_1$ with $\bar{\sigma}(\Delta_1) \leq \beta$, the perturbed closed-loop system is well-posed and stable if and only if

$$\sup_{\mathrm{Re}(s) \geq 0} \mu_1(P_{11}(s)) < \frac{1}{\beta}.$$

(2) For all $\Delta_1 \in \mathbf{\Delta}_1$ with $\bar{\sigma}(\Delta_1) \leq \beta$, the perturbed closed-loop system is well-posed,

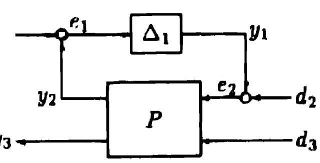


FIG. 9. Uncertain system for robustness tests.

stable and

$$\sup_{\mathrm{Re}(s)\geq 0} \mu_2[\mathcal{S}(\Delta_1, P(s))] < \frac{1}{\beta},$$

if and only if

$$\sup_{\mathrm{Re}(s)\geq 0} \mu_{\mathbf{\Delta}}(P(s)) < \frac{1}{\beta}.$$

Proof. The proof follows from the definitions of μ, well-posedness, stability, and invertibility of matrices with elements in a commutative ring.

Remark 5.2. Although the structured singular value is not necessarily a norm, we introduce the following notation: for a proper, rational matrix P, analytic in the closed-right-half-plane, and block structure $\mathbf{\Delta}$ of appropriate dimension, define

$$\|P\|_{\mathbf{\Delta}} := \sup_{\mathrm{Re}(s)\geq 0} \mu_{\mathbf{\Delta}}(P(s)).$$

Remark 5.3. In Section 6, techniques that allow the right-half plane supremums to be replaced (equivalently) by imaginary-axis supremums will be developed.

It is possible to easily generalize these robustness theorems to the case where Δ is a block-diagonal, finite dimensional, stable, linear time-invariant system (as opposed to a constant, complex matrix). Let $\mathbf{\Delta}$ be a block structure, as in equation (3.1). We want to consider feedback perturbations to P which are themselves dynamical systems, with the block-diagonal structure of the set $\mathbf{\Delta}$. To do so, first let $\mathcal{M}(\mathbf{S})$ denote the set of rational, proper, stable, transfer matrices. Associated with any block structure $\mathbf{\Delta}$, let $\mathcal{M}(\mathbf{\Delta})$ denote the set of all block diagonal, stable rational transfer functions, with block structure like $\mathbf{\Delta}$.

$$\mathcal{M}(\mathbf{\Delta}) := \{\Delta(\cdot) \in \mathcal{M}(\mathbf{S}) : \Delta(s_0) \in \mathbf{\Delta} \text{ for all } s_0 \in \bar{\mathbf{C}}_+\}.$$

For any $\Delta_1 \in \mathcal{M}(\mathbf{\Delta}_1)$, the closed-loop system is said to be:

- well-posed if $\det(I - P_{11}(\infty)\Delta_1(\infty)) \neq 0$. This is the necessary and sufficient condition that all closed-loop transfer functions in Fig. 5.1 be proper;
- stable if all closed-loop transfer functions in Fig. 5.1 are analytic in the closed right-half-plane.

Theorem 5.4. Suppose that $P(s)$ has all of its poles in the open left-half plane. Let $\beta > 0$. Then:

(1) For all $\Delta_1 \in \mathcal{M}(\mathbf{\Delta}_1)$ with $\|\Delta_1\|_\infty \leq \beta$, the perturbed closed-loop system is well-posed and stable if and only if

$$\sup_{\mathrm{Re}(s)\leq 0} \mu_1(P_{11}(s)) < \frac{1}{\beta}.$$

(2) For all $\Delta_1 \in \mathcal{M}(\mathbf{\Delta}_1)$ with $\|\Delta_1\|_\infty \leq \beta$, the perturbed closed-loop system is well-posed, stable and

$$\sup_{\mathrm{Re}(s)\geq 0} \mu_2[\mathcal{S}(\Delta_1(s), P(s))] < \frac{1}{\beta},$$

if and only if

$$\sup_{\mathrm{Re}(s)\geq 0} \mu_{\mathbf{\Delta}}(P(s)) < \frac{1}{\beta}.$$

In summary, the peak value on the μ plot of the frequency response, that the perturbation "sees", determines the size of perturbations that the loop is robustly stable (and/or performing) against.

Other, more sophisticated assumptions about the perturbations may be formulated, and solved with μ. These include gap/graph topology uncertainty (Foo and Postlethwaite, 1988; Khargonekar and Kaminer, 1991), and different induced norms to measure the size of the uncertainty, (Bamieh and Dahleh, 1992).

6 MAXIMUM-MODULUS THEOREM FOR LFTS WITH μ

This section describes a maximum-modulus theorem that μ satisfies: μ of an LFT on a norm-bounded structured set achieves its maximum on the unitary elements of this set. This is a generalization of the ordinary maximum-modulus theorem for rational functions of a complex variable. We begin by stating a well-known result from complex analysis namely that the roots of a polynomial are continuous functions of the coefficients of the polynomial.

Lemma 6.1. Let $f(z) = \sum_{i=0}^{n} a_i z^i$ be an nth order polynomial, $a_n \neq 0$. Let $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_n$ be the roots of f. For any $\epsilon > 0$ and any integer $m > 0$, there exists a $\delta_{m,\epsilon} > 0$ such that if $g(z)$, defined by

$$g(z) = \sum_{i=0}^{m} b_i z^i,$$

has coefficients $b_i \in \mathbf{C}$ which satisfy $|b_i| < \delta_{m,\epsilon}$, then there are n roots of $f + g$, labeled $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n$ that satisfy $|\tilde{z}_i - \bar{z}_i| < \epsilon$.

Next, we shift our attention to polymomials in several dimensions, that is, polynomials taking $\mathbf{C}^k \to \mathbf{C}$. If $z \in \mathbf{C}^k$, we let $\|z\|_\infty := \max_{i\leq k} |z_i|$. For

$p: \mathbf{C}^k \to \mathbf{C}$, a polynomial, define β_p as

$$\beta_p = \min \{\|z\|_\infty : p(z) = 0\}. \qquad (6.1)$$

In other words, β_p is the norm of the minimum norm roots of the polynomial. The next two lemmas are from Doyle (1982). The first concerns minimum norm roots and is a direct consequence of Lemma 6.1. The second provides the essential argument of the maximum-modulus theorem for μ.

Lemma 6.2. Let p be a polynomial from $\mathbf{C}^k \to \mathbf{C}$. Define β_p via (6.1). Then there exists a $z \in \mathbf{C}^k$ such that $|z_i| = \beta_p$ for each i, and $p(z) = 0$.

Sketch of proof. A proof is given in Doyle (1982). The main idea is as follows: let $\bar{z} \in \mathbf{C}^k$ be a root of p with $\beta_p = \|\bar{z}\|_\infty$. If all of the coordinates of $\bar{z}$ satisfy $|\bar{z}_i| = \beta_p$, then stop. Otherwise, take one of the coordinates of $\bar{z}$, say $\bar{z}_1$ whose magnitude is less than β_p. Consider the polynomial $q(z_1) := p(z_1, \bar{z}_2, \ldots, \bar{z}_k)$. This has a root at $\bar{z}_1$, and $|\bar{z}_1| < \beta_p$. If this is a nontrivial polynomial, then, by slightly reducing (in magnitude) all of the $\bar{z}_i$, $i \geq 2$, the coefficients of the polynomial change slightly, and it has a root very close to $\bar{z}_1$. This implies that p has a root $\tilde{z} \in \mathbf{C}^k$ such that $\|\tilde{z}\|_\infty < \beta_p$, which contradicts the definition. On the other hand, if the polynomial $q \equiv 0$, then the variable z_1 does not matter, and we can repeat the argument on a different coordinate, say z_2, that satisfies $|\bar{z}_2| < \beta_p$.

Lemma 6.3. Let $\mathbf{\Delta} \subset \mathbf{C}^{n \times n}$ be a given block structure, and let $\mathbf{Q}$ be defined as in Section 3. If $M \in \mathbf{C}^{n \times n}$ has $\mu_\Delta(M) = 1$, then there is a $Q \in \mathbf{Q}$ such that $\det(I - MQ) = 0$.

Proof. Since $\mu_\Delta(M) = 1$, there is a $\hat{\Delta} \in \mathbf{\Delta}$ with $\bar{\sigma}(\hat{\Delta}) = 1$ and $\det(I - M\hat{\Delta}) = 0$. Also, for any $\Delta \in \mathbf{\Delta}$ with $\bar{\sigma}(\Delta) < 1$, the matrix $I - M\Delta$ is nonsingular. Do a singular value decomposition on each block that makes up $\hat{\Delta}$. This gives $U, V \in \mathbf{Q}$, and a diagonal $\hat{\Sigma} \in \mathbf{\Delta}$, such that

$$\det(I + MU\hat{\Sigma}V^*) = 0.$$

Since $\hat{\Sigma} \in \mathbf{\Delta}$ is diagonal, it appears as

$$\hat{\Sigma} = \operatorname{diag}[\hat{\delta}_1 I_{r_1}, \ldots, \hat{\delta}_S I_{r_S}, \hat{\alpha}_1, \ldots, \hat{\alpha}_w],$$

for some nonnegative real numbers $\hat{\delta}_i$ and $\hat{\alpha}_j$, and $w = \sum_{j=1}^{E} m_j$ (recall that the jth full block is $m_j \times m_j$, hence each full block contributes m_j of the α's). With $\bar{\sigma}(\hat{\Delta}) = 1$, at least one of the $\hat{\delta}_i$ or $\hat{\alpha}_j$ is 1.

Consider $S + w$ complex variables, $z_1, \ldots, z_{S+w}$. Define a variable Σ by

$$\Sigma = \operatorname{diag}[z_1 I_{r_1}, \ldots, z_S I_{r_S}, z_{S+1}, \ldots, z_{S+w}].$$

Then $\det(I + MU\Sigma V^*)$ is a polynomial on $\mathbf{C}^{S+w}$, since the determinant involves only multiplications and additions of its argument. Since $\mu_\Delta(M) = 1$, a minimum norm (using $\|\cdot\|_\infty$ on $\mathbf{C}^{S+w}$, as above) root of this polynomial has norm equal to 1. Let $\bar{\Sigma}$ be the particular minimizing root with all components of equal magnitude, namely one. Then we can write $\bar{\Sigma} = \Phi$ for some $\Phi \in \mathbf{Q}$. This gives

$$\det(I + MU\Phi V^*) = 0.$$

Defining $Q := U\Phi V^*$ completes the proof.

The next theorem from Doyle (1982) follows immediately from Lemma 6.3 and the facts that $Q \in \mathbf{Q}$ implies $\rho(QM) \leq \mu_\Delta(M)$ and $\det(I - MQ) = 0$ implies that $\rho(QM) \geq 1$.

Theorem 6.4.

$$\max_{Q \in \mathbf{Q}} \rho(QM) = \max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M) = \mu_\Delta(M).$$

Hence the lower bound given for μ in equation (3.10) is actually not just a bound, but an equality.

To motivate the main result of the section, recall the general setup for the linear fractional transformation $\mathcal{S}(M, \Delta)$. Suppose $M \in \mathbf{C}^{(n_1+n_2)\times(n_1+n_2)}$ is given. We partition it in the obvious way

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \qquad (6.2)$$

with $M_{ij} \in \mathbf{C}^{n_i \times n_j}$. Let $\mathbf{\Delta}_1 \subset \mathbf{C}^{n_1 \times n_1}$ and $\mathbf{\Delta}_2 \subset \mathbf{C}^{n_2 \times n_2}$ be two block structures and define $\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{Q}_1$, and $\mathbf{Q}_2$ correspondingly.

The maximum-modulus theorem from Packard and Balsamo (1988) is as follows.

Theorem 6.5. Let M be given as in (6.2), along with two block structures $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. Suppose that $\mu_2(M_{22}) < 1$. Then

$$\max_{Q_2 \in \mathbf{Q}_2} \mu_1(\mathcal{S}(M, Q_2)) = \max_{\Delta_2 \in \mathbf{B}_2} \mu_1(\mathcal{S}(M, \Delta_2)). \qquad (6.3)$$

Proof. A detailed proof can be found in the reference. The main idea is as follows: suppose (by an appropriate scaling) that the maximum on the right hand side of equation (6.3) is one. Then, since $\mu_2(M_{22}) < 1$, it is possible to show that $\mu_\Delta(M) = 1$. Using Lemma 6.3, construct matrices Q_1 and Q_2 such that $I - M \operatorname{diag}[Q_1, Q_2]$ is singular. Again, use the fact that $\mu_2(M_{22}) < 1$ to conclude that $I -$

$\mathcal{S}(M, Q_2)Q_1$ is singular. This shows that $\mu_1(\mathcal{S}(M, Q_2)) \ge 1$, completing the argument.

Remark 6.6. This is similar to a result in Boyd and Desoer (1985): that for functions $H(z)$ analytic on the disk, the function $\mu(H(z))$ achieves its maximum on the boundary: $\max_{|z|\le 1} \mu(H(z)) = \max_{|z|=1} \mu(H(z))$. It is possible to use their result to derive Theorem 6.5 and vice versa.

It is instructive to see how Theorem 6.4 can be obtained as a special case of Theorem 6.5. Let $\mathbf{\Delta} \subset \mathbf{C}^{n\times n}$ be a given block structure, with associated sets $\mathbf{B}_\Delta$ and $\mathbf{Q}$. Define $\mathbf{\Delta}_1 := \{\delta I_n : \delta \in \mathbf{C}\}$, and for each $M \in \mathbf{C}^{n\times n}$, define $\bar{M}$ as

$$\bar{M} := \begin{bmatrix} 0 & M \\ I_n & 0 \end{bmatrix}.$$

By Theorem 6.5, and noting that $\mu_1(\cdot) = \rho(\cdot)$, we have

$$\begin{aligned} \max_{Q\in\mathbf{Q}} \rho(MQ) &= \max_{Q\in\mathbf{Q}} \mu_1(\mathcal{S}(\bar{M}, Q)) \\ &= \max_{\Delta\in\mathbf{B}_\Delta} \mu_1(\mathcal{S}(\bar{M}, \Delta)) \\ &= \max_{\Delta\in\mathbf{B}_\Delta} \rho(M\Delta) \\ &= \mu_\Delta(M). \end{aligned} \tag{6.4}$$

This is exactly Theorem 6.4.

7. LOWER BOUND POWER ALGORITHM

This section presents an iterative algorithm to compute lower bounds for the structured singular value. The algorithm resembles a mixture of power methods for eigenvalues and singular values, which is not surprising, since the structured singular value can be viewed as a generalization of both. If the algorithm converges, a lower bound for μ results. We prove that μ is always an equilibrium point of the algorithm.

In Fan and Tits (1986) the calculation of μ is reformulated as a smooth optimization problem. As with all of the known exact expressions for μ, the function to be maximized has local maximums which are not global, so in general the method yields only lower bounds for μ. Similar comments can be made for the ideas in Doyle (1982) and Helton (1988), as well as the algorithm in this section. The contribution here is yet another lower bound algorithm to aid in the analysis of robustness of systems with structured uncertainty, along with a deeper conceptual understanding of the structured singular value.

We begin by noting that both the functions $r: \mathbf{B}_\Delta \to \mathbf{R}$, defined by $r(\Delta) := \rho(\Delta M)$ and $\bar{r}: \mathbf{Q} \to \mathbf{R}$, defined by $\bar{r}(Q) := \rho(QM)$ have local maximums which are not global. Note, though, that the function $\bar{r}$ is a restriction of r, and it is possible to construct examples where a point $Q \in \mathbf{Q}$ is a local maximum of $\bar{r}$, but not a local maximum of r. Such a point definitely does not correspond to the maximizer that gives $\mu_\Delta(M)$, and so we will not consider the corresponding lower bound from such points as acceptable. Rather, acceptable lower bounds will correspond to points $Q \in \mathbf{Q}$ which are local maximums of the function r.

Roughly speaking, this section develops an iterative algorithm which ultimately generates a point $Q \in \mathbf{Q}$ that is a local maximum of the function r. In general, these are a proper subset of the local maximums of the function $\bar{r}$, though the global maximums over the two sets are the same. Some of the preliminary results are generalizations of those found in Fan and Tits (1986) and Daniel *et al.* (1986).

We will be interested in local maximums of the function $r(\Delta) = \rho(\Delta M)$, therefore we begin with some facts from perturbation theory, which assist in characterizing local maximums.

7.1. *Matrix facts*

In this section, we collect a few useful facts.

Suppose $W: \mathbf{R} \to \mathbf{C}^{n\times n}$ is an analytic function of the real parameter t. If λ_0 is an eigenvalue of $W_p := W(0)$ of multiplicity one, then for some open interval containing 0, this eigenvalue is an analytic function of t, as are the eigenvectors associated with it. That is, suppose there are $x_0, y_0 \in \mathbf{C}^n$, satisfying $y_0^* x_0 = 1$, $W_0 x_0 = \lambda_0 x_0$, and $y_0^* W = \lambda_0 y_0^*$. Then there is an $\epsilon > 0$ and analytic functions $x: (-\epsilon, \epsilon) \to \mathbf{C}^n$, $y: (-\epsilon, \epsilon) \to \mathbf{C}^n$, and $\lambda: (-\epsilon, \epsilon) \to \mathbf{C}$, such that $x(0) = x_0$, $y(0) = y_0$, $\lambda(0) = \lambda_0$ and for all $t \in (-\epsilon, \epsilon)$

$$\begin{aligned} y^* x &= 1, \\ Wx &= \lambda x, \\ y^* W &= \lambda y^*. \end{aligned} \tag{7.1}$$

This follows from Kato (1982). We can then differentiate and obtain $\dot{\lambda}(0) = y_0^* \dot{W}(0) x_0$.

The next two lemmas follow from elementary linear algebra. They will be used in the main theorem of the next section.

Lemma 7.1. Let $y, x \in \mathbf{C}^n$ with $y \ne 0$ and $x \ne 0$. There exists $d \in \mathbf{R}$, $d > 0$, such that $(1/\sqrt{d})y = \sqrt{d}x$ if and only if $\mathrm{Re}\,(y^* G x) \le 0$ for every $G \in \mathbf{C}^{n\times n}$ satisfying $G + G^* \le 0$.

Lemma 7.2. Let x and y be two nonzero vectors

in $\mathbf{C}^n$. Then there exists a Hermitian, positive definite $D \in \mathbf{C}^{n \times n}$, such that $D^{1/2}x = D^{-1/2}y$ if and only if $\mathrm{Re}\,(gy^*x) \le 0$ for every $g \in \mathbf{C}$ with $g + \bar{g} \le 0$.

The condition in Lemma 7.2 on y^*x involving $g \in \mathbf{C}$ is equivalent to y^*x being real and positive. We have chosen to write it in the form above so that it is a natural analog to Lemma 7.1 and is stated exactly as it will be applied in Theorem 7.3 of the next section.

7.2. *Eigenvector characterization of local maximums*

Consider the function $r: \mathbf{B}_\Delta \to \mathbf{R}$, defined by $r(\Delta) = \rho(\Delta M)$. Recall that $\mu_\Delta(M) = \max_{\Delta \in \mathbf{B}_\Delta} r(\Delta)$, and that the global maximum occurs on the subset $\mathbf{Q} \subset \mathbf{B}_\Delta$. In this section, we characterize the occurrence of a local maximum of r at $\Delta = I \in \mathbf{Q} \subset \mathbf{B}_\Delta$, in terms of the eigenvectors of M. We begin with some notation.

Let x and y be nonzero right and left eigenvectors of M, associated with an eigenvalue λ: $Mx = \lambda x$ and $y^*M = \lambda y^*$. Partition x and y compatibly with the block structure Δ,

$$x = \begin{bmatrix} x_{r_1} \\ x_{r_2} \\ \vdots \\ x_{r_S} \\ x_{m_1} \\ x_{m_2} \\ \vdots \\ x_{m_F} \end{bmatrix}, \quad y = \begin{bmatrix} y_{r_1} \\ y_{r_2} \\ \vdots \\ y_{r_S} \\ y_{m_1} \\ y_{m_2} \\ \vdots \\ y_{m_F} \end{bmatrix} \qquad (7.2)$$

where $x_{r_i}, y_{r_i} \in \mathbf{C}^{r_i}$ and $x_{m_j}, y_{m_j} \in \mathbf{C}^{m_j}$ for each i and j. We call these the "block components" of x and y, and for technical reasons, we define a nondegeneracy condition: x and y are nondegenerate if for every i, $y_{r_i}^* x_{r_i} \ne 0$, and for each j, $x_{m_j} \ne 0$, $y_{m_j} \ne 0$. We will also assume that $\rho(M) = \lambda_0$ is a distinct eigenvalue of M.

The condition that $\rho(M) = \lambda_0 > 0$ is without loss of generality, because Δ can always be used to enforce this (for any ϕ, $e^{j\phi}\Delta = \Delta$). The conditions of nondegeneracy and λ_0 distinct are not so easily dispensed with and there are basically two approaches to deal with them. The first would be to argue that there are generic conditions and thus unlikely to cause problems in practice. A far more satisfactory solution is to generalize the theorems and proofs in this section to remove them. In fact, this can be done, but not without substantial additional technical complication. Since the results in this subsection are presented primarily to give insight into the power algorithms to be presented in the next subsection, these additional technicalities have been foregone in favor of a simpler development.

Theorem 7.3. Let $M \in \mathbf{C}^{n \times n}$ be given, and suppose $\lambda_0 > 0$ is a distinct eigenvalue of M, with nondegenerate right and left eigenvectors x and y. Suppose that $\rho(M) = \lambda_0$. If the function $r: \mathbf{B}_\Delta \to \mathbf{R}$ defined by $r(\Delta) = \rho(\Delta M)$ has a local maximum (with respect to the set $\mathbf{B}_\Delta$) at $\Delta = I$, then there exists a $D \in \mathbf{D}$ such that $D^{-1/2}y = D^{1/2}x$.

Proof. Let $G \in \Delta$ with $G + G^* \le 0$ so that G has the form

$$\mathrm{diag}\,[g_1 I_{r_1}, \ldots, g_S I_{r_S}, G_1, \ldots, G_F], \qquad (7.3)$$

where $\mathrm{Re}\,(g_i) \le 0$, and $G_j + G_j^* \le 0$ for all i and j and $e^{Gt} \in \mathbf{B}_\Delta$ for all $t \ge 0$ with $e^{Gt} = I$ for $t = 0$. Define a matrix function $W: \mathbf{R} \to \mathbf{C}^{n \times n}$ by $W(t) := e^{Gt}M$. Note that at $t = 0$, λ_0 is a simple eigenvalue of $W(0)$, with x and y the right and left eigenvectors. For some nonempty interval containing 0, this eigenvalue is always simple, and hence there is an analytic function of the real variable t, $\lambda(t)$, defined on that interval, such that $\lambda(t)$ is an eigenvalue of $W(t)$ for all t and $\lambda(0) = \lambda_0$. It is easy to calculate $\dot\lambda(0)$, namely

$$\dot\lambda(0) = y^* \dot W(0) x = \lambda_0 y^* G x. \qquad (7.4)$$

By hypothesis, $\lambda_0 > 0$, $\rho(M) = \lambda_0$ and the function $\rho(\Delta M)$ has a local maximum (with respect to $\mathbf{B}_\Delta$) at $\Delta = I$. Therefore

$$\mathrm{Re}\left(\frac{d}{dt}\lambda(t)\bigg|_{t=0}\right) \le 0, \qquad (7.5)$$

which says that the magnitude of λ must be nonincreasing at $t = 0$. Using the "block notation" of (7.2) and substituting (7.3) and (7.4) into (7.5) yields

$$\mathrm{Re}\left(\sum_{i=1}^{S} g_i y_{r_i}^* x_{r_i} + \sum_{j=1}^{F} y_{m_j}^* G_j x_{m_j}\right) \le 0. \qquad (7.6)$$

This must hold for arbitrary $G \in \Delta$ satisfying $G + G^* \le 0$. Applying Lemmas 7.1 and 7.2 we conclude that for each i, there is a $D_i = D_i^* \in \mathbf{C}^{n \times n}$, $D_i > 0$ such that $D_i^{-1/2} y_{r_i} = D_i^{1/2} x_{r_i}$, and for each j, there is a $d_j \in \mathbf{R}$, $d_j > 0$ such that $1/\sqrt{d_j}\, y_{m_j} = \sqrt{d_j}\, x_{m_j}$. Arranging all of these D_is and d_js into one block diagonal D completes the proof.

Remark 7.4. Note that assuming λ_0 is distinct assures differentiability (Kato, 1982). Since λ_0 is a solution of $\max_{\Delta \in \mathbf{B}_\Delta} \max_i |\lambda_i(\Delta M)|$, it is likely that at the maximum it will be distinct. In any case, if λ_0 is not distinct, it can still be shown to be differentiable at a local maximum, and the rest

of the proof remains. Unfortunately, this proof of differentiability is tedious and technical and for this reason has been omitted.

Theorem 7.5. Let $Q_0 \in \mathbf{Q}$ achieve the global optimum for the problem $\max_{Q \in \mathbf{Q}} \rho(QM)$. Suppose that the eigenvalue associated with $\rho(Q_0 M)$ is distinct, real and positive, and hence equal to $\mu = \mu_\Delta(M)$. If x and y are nondegenerate right and left eigenvectors of the eigenvalue μ, then there exists a $D \in \mathbf{D}$, and $\xi \in \mathbf{C}^n$, $\|\xi\| = 1$ such that

$$\begin{aligned} Q_0 D^{1/2} M D^{-1/2} \xi &= \mu \xi, \\ \xi^* Q_0 D^{1/2} M D^{-1/2} &= \mu \xi^*. \end{aligned} \tag{7.7}$$

Proof. By Theorem 6.4, any global maximizer of $\max_{Q \in \mathbf{Q}} \rho(QM)$, is also a maximizer of $\max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta M)$. Define $\tilde{M} := Q_0 M$, then $\Delta = I$ is a local (in fact global) maximizer for $\max_{\Delta \in \mathbf{B}_\Delta} \rho(\Delta \tilde{M})$. Apply Theorem 7.3 to the matrix $\tilde{M}$ and define $\xi = D^{1/2} x = D^{-1/2} y$ to prove the theorem.

Remark 7.6. This result was first shown in Fan and Tits (1986) for the case of $S = 0$. It is also similar to the "principal direction alignment" ideas in Daniel *et al.* (1986). Theorem 7.5 is more general, though, since it handles repeated scalar blocks as well as full blocks.

Remark 7.7. This theorem is not true if we consider local maximums that are not global of the function $\tilde{r}: \mathbf{Q} \to \mathbf{R}$ defined as $\tilde{r}(Q) := \rho(QM)$.

Remark 7.8. Any real number $\beta > 0$ satisfying

$$\begin{aligned} Q D^{1/2} M D^{-12} \xi &= \beta \xi, \\ \xi^* Q D^{1/2} M D^{-1/2} &= \beta \xi^*, \end{aligned} \tag{7.8}$$

for some $Q \in \mathbf{Q}$, $D \in \mathbf{D}$ and nonzero $\xi \in \mathbf{C}^n$ is a lower bound for $\mu_\Delta(M)$. This follows because $I - (1/\beta) QM$ is a singular matrix.

7.3. *Lower bound power algorithm*

In this section, we propose an iterative algorithm (reminiscent of the power algorithm for spectral radius) to find solutions to equations (7.7) and therefore get lower bounds for μ.

Rewriting (7.7), and changing notation a bit, we want to find $Q \in \mathbf{Q}$, $D \in \mathbf{D}$, $\beta > 0$, and $\xi \in \mathbf{C}^n$ with $\|\xi\| = 1$ such that

$$\begin{aligned} Q D^{1/2} M D^{-1/2} \xi &= \beta \xi, \\ D^{-1/2} M^* D^{1/2} Q^* \xi &= \beta \xi. \end{aligned}$$

These two constraint equations can be rewritten as

$$\begin{aligned} M(D^{-1/2}\xi) &= \beta(D^{-1/2} Q^* \xi), \\ M^*(D^{-1/2} Q^* \xi) &= \beta(D^{1/2} \xi). \end{aligned}$$

For a given D, Q, and ξ, define vectors a, b, z, and w by

$$\begin{aligned} b &:= D^{-1/2}\xi, & a &:= D^{-1/2} Q^* \xi, \\ z &:= D^{1/2} Q^* \xi, & w &:= D^{1/2} \xi. \end{aligned} \tag{7.9}$$

With this definition, we have $Mb = \beta a$ and $M^* z = \beta w$. We can eliminate ξ from (7.9) to get

$$\begin{aligned} b &= Qa = D^{-1} w, \\ z &= Da = Q^* w. \end{aligned} \tag{7.10}$$

Since the unknowns Q and D generally may have high dimension, we would like to write the four relationships from equation (7.10) in a manner that does not involve the matrices Q and D. With a few technical conditions, this can be done. In order to simplify the upcoming formulas, we will consider a block structure with $S = 1$, $F = 1$ (by duplicating the appropriate formulas for additional blocks, whether they are repeated scalar blocks or full blocks, it is straightforward to extend the algorithm to more general structures). Hence the sets $\mathbf{D}$ and $\mathbf{Q}$ are

$$\begin{aligned} \mathbf{D} = \{ &\operatorname{diag}[D_1, d_2 I_{m_1}] : D_1 \in \mathbf{C}^{r_1 \times r_1}, \\ &D_1 = D_1^* > 0, d_2 > 0\}, \end{aligned} \tag{7.11}$$

$$\begin{aligned} \mathbf{Q} = \{ &\operatorname{diag}[q_1 I_{r_1}, Q_2] : \bar{q}_1 q_1 = 1, \\ &Q_2 \in \mathbf{C}^{m_1 \times m_1}, Q_2^* Q_2 = I_{m_1}\}. \end{aligned} \tag{7.12}$$

With respect to this, we will partition the vectors accordingly, so $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$, where $z_1 \in \mathbf{C}^{r_1}$ and $z_2 \in \mathbf{C}^{m_1}$, and likewise for the other vectors.

Lemma 7.9. Let r_1 and m_1 be positive integers. Let z_1, w_1, b_1, $a_1 \in \mathbf{C}^{r_1}$ and z_2, w_2, b_2, $a_2 \in \mathbf{C}^{m_1}$ be nonzero vectors with $a_1^* w_1 \neq 0$. Then, there exists a $D \in \mathbf{D}$ and $Q \in \mathbf{Q}$ such that

$$\begin{aligned} b &= Qa, & z &= Q^* w, \\ z &= Da, & b &= D^{-1} w, \end{aligned}$$

if and only if

$$\begin{aligned} z_1 &= \frac{w_1^* a_1}{|w_1^* a_1|} w_1, & z_2 &= \frac{\|w_2\|}{\|a_2\|} a_2, \\ b_1 &= \frac{a_1^* w_1}{|a_1^* w_1|} a_1, & b_2 &= \frac{\|a_2\|}{\|w_2\|} w_2. \end{aligned}$$

Proof.

The relations for z_2 and b_2 follow by direct substitution. For z_1 and b_1, it is easiest to

define an auxiliary variable $\zeta := D^{1/2}b$, and then verify via substitutions.

$\leftarrow$ Let $q_1 = a_1^* w_1 / |a_1^* w_1|$, since this is well defined. Likewise, choose $d_2 = \|w_2\| / \|a_2\|$. By assumption, d_2 is well defined, and nonzero. Since $\|w_2\| = \|z_2\|$, let Q_2 be any unitary matrix that takes w_2 into z_2. The matrix Q_2 also rotates b_2 into a_2,

$$Q_2 b_2 = \frac{1}{d_2} Q_2 w_2 = \frac{1}{d_2} z_2 = a_2.$$

Next, we calculate $a_1^* z_1 = |a_1^* w_1|$, which is nonzero by assumption: hence Lemma 7.2 yields a Hermitian, positive definite D_1 such that $D_1 a_1 = z_1$. As we hope, D_1 takes b_1 into w_1 too,

$$D_1 b_1 = q_1 D_1 a_1 = q_1 z_1 = w_1.$$

Defining D and Q in the obvious manner completes the proof.

We are now prepared for the main theorem.

Theorem 7.10. Let $M \in \mathbf{C}^{n\times n}$ be given, and let $\mathbf{\Delta}$ be the two block ($S = 1$, $F = 1$) structure defined above, with block sizes r_1 and m_1, where $r_1 + m_1 = n$. Suppose $\beta > 0$ is given. Then there exists $Q \in \mathbf{Q}$, $D \in \mathbf{D}$, $\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \in \mathbf{C}^n$, $\|\xi\| = 1$, $\xi_1 \neq 0$, $\xi_2 \neq 0$ with

$$\begin{aligned} QD^{1/2}MD^{-1/2}\xi &= \beta\xi, \\ \xi^* QD^{1/2}MD^{-1/2} &= \beta\xi^*, \end{aligned} \tag{7.13}$$

if and only if there exists nonzero vectors a_1, w_1, b_1, $a_1 \in \mathbf{C}^{r_1}$ and z_2, w_2, b_2, $a_2 \in \mathbf{C}^{m_1}$ with $a_1^* w_1 \neq 0$ and

$$\begin{aligned} \beta a &= Mb, \\ z_1 &= \frac{w_1^* a_1}{|w_1^* a_1|} w_1, \quad z_2 = \frac{\|w_2\|}{\|a_2\|} a_2, \\ \beta \omega &= M^* z, \\ b_1 &= \frac{a_1^* w_1}{|a_1^* w_1|} a_1, \quad b_2 = \frac{\|a_2\|}{\|w_2\|} w_2. \end{aligned} \tag{7.14}$$

Remark. In order to find decompositions using the representation that this theorem allows (equation (7.14)—free of Qs and Ds), we can restrict ourselves to unit vectors a, b, z, w. Why? Suppose there are nonzero vectors satisfying (7.14). Examining the equations, it is clear that scaling z and w by some $\alpha \neq 0$ and scaling b and a by some $\gamma \neq 0$ does not affect any of the equalities in (7.14). Moreover, the equalities in (7.14) always imply that $\|z\| = \|w\|$, and $\|a\| = \|b\|$, so by proper scaling, all the vectors would be unit norm.

In the above theorem, we have purposefully written the conditions (7.14) in a manner that suggests attempting to find a solution in an iterative fashion. In particular, for $i = 1, 2$, let vectors a_{i_k}, b_{i_k}, z_{i_k}, and w_{i_k}, and positive scalars $\tilde{\beta}_k$, $\hat{\beta}_k$ evolve as

$$\begin{aligned} \tilde{\beta}_{k+1} a_{k+1} &= Mb_k, \\ z_{1_{k+1}} &= \frac{w_{1_k}^* a_{1_{k+1}}}{|w_{1_k}^* a_{1_{k+1}}|} w_{1_k}, \\ z_{2_{k+1}} &= \frac{\|w_{2_k}\|}{\|a_{2_{k+1}}\|} a_{2_{k+1}}, \\ \hat{\beta}_{k+1} w_{k+1} &= M^* z_{k+1}, \\ b_{1_{k+1}} &= \frac{a_{1_{k+1}}^* w_{1_k}}{|a_{1_{k+1}}^* w_{1_{k+1}}|} a_{1_{k+1}}, \\ b_{2_{k+1}} &= \frac{\|a_{2_{k+1}}\|}{\|w_{2_{k+1}}\|} w_{2_{k+1}}, \end{aligned} \tag{7.15}$$

where the values of $\tilde{\beta}_{k+1}$ and $\hat{\beta}_{k+1}$ are chosen >0, so that $\|a_{k+1}\| = \|w_{k+1}\| = 1$.

Note also that if the initial b and w vectors that start the iteration are unit vectors, then at every step, all vectors, a, b, z, and w will be unit length.

Remarks.

(*a*) Potential problems within the iteration are:

- $Mb_k = 0$ or $M^* z_k = 0$, then a_{k+1} or w_{k+1} is not well defined.
- $a_{1_k}^* w_{1_k} = 0$, then the vectors $a_{1_{k+1}}$ and/or $b_{1_{k+1}}$ are not well defined.
- Either $\|w_{2_k}\| = 0$ or $\|a_{2_k}\| = 0$, making b_{2_k} and/or z_{2_k} not well defined.

If any of these conditions occur, then one possibility is to restart the algorithm at a different initial condition (i.e. a new b_{1_0}, b_{2_0}, w_{1_0} and w_{2_0}). A more sophisticated approach is to examine the above conditions and recognize that a sensible iteration can still be defined even if these conditions occur. Algorithms have been developed along these lines and will be discussed elsewhere.

(b) If the iteration does converge to an equilibrium point, then the β values must be equal, that is $\tilde{\beta} = \hat{\beta}$. This is easy to see: suppose the equations in (7.14) are satisfied (convergence of the algorithm in (7.15)) but the β associated with b and a is $\tilde{\beta}$ and the β associated with z and w is $\hat{\beta}$. The converged equations imply that there exists a $Q \in \mathbf{Q}$ and $D \in \mathbf{D}$ such that $QD^{1/2}MD^{-1/2}\ (D^{1/2}b) = \tilde{\beta}(D^{1/2}b)$ and $(QD^{1/2}MD^{-1/2})^*(D^{1/2}b) = \hat{\beta}(D^{1/2}b)$. Since the βs are real, they must be equal. Hence, when verifying convergence of the algorithm, it is necessary to begin checking the

convergence of the vectors only after the $\tilde{\beta}_k$ and $\hat{\beta}_k$ values are nearly equal. This saves some computations early in the iteration.

(c) • If there were only the first block, which is a repeated scalar block, the iteration would be a power iteration for the largest (in magnitude) eigenvalue of the matrix M. Since μ for one repeated scalar block is the spectral radius, the algorithm we have proposed reduces to a valid algorithm in the special case of one repeated scalar block.

• If there were only the second block, which is a full block, the iteration becomes a eigenvalue power algorithm for M^*M, hence it will give the largest singular value of M. Again, with respect to this specific block structure, this is what we want.

Hence, the iteration we have proposed is a mix of two separate, well understood iterations, both of which converge to the largest eigenvalue/singular value. We might hope that this algorithm will converge to the largest β for which the equations in (7.8) are solved, which by Theorem 7.5 is equal to $\mu_\Delta(M)$. Unfortunately, this is not always the case.

Extensive computational experience (Balas *et al.*, 1991; Packard *et al.*, 1988) has led to the following conclusions.

(1) The algorithm works well in practice, and versions of it have been used very extensively in universities and industry. It appears to have roughly order n^2 growth rates for computation as a function of problem size. The main difficulty is that it occasionally does not converge or converges to a value of β which is not μ.

(2) The difficulties described above do not seem to occur in practice, however there are matrices whose optimally scaled eigenvector block components (equation (7.2)) do not satisfy the "nonzero" block conditions described at the beginning of Section 7.2. This type of situation will lead to the difficulties mentioned. In any event, while it is easy to construct matrices where these problems happen, running the algorithm on frequency responses of actual closed loop systems has not been a problem.

(3) Limit cycles can occur, and seem to occur more often when there are large repeated scalar blocks. Unlike an equilibrium point, the presence of a stable limit cycle does not immediately give rise to a lower bound for μ.

(4) In general, there are several stable equilibrium points, with different values of β. This is in contrast with the conventional power algorithms for ρ and $\bar{\sigma}$, where only the largest ones are stable. It is even possible that the algorithm converges to a value of β which is smaller than $\rho(M)$.

(5) It is possible to refine the power algorithm to guarantee convergence to some local maximum, but at the expense of greater computation time. We are currently researching algorithms that give favorable tradeoffs between convergence properties and running times.

8. RELATING μ AND $\inf_{D\in\mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$

The purpose of this section is to study the relationship between $\mu_\Delta(M)$ and the upper bound. The two-step strategy we take first involves characterizing the optimality conditions for the upper bound, and then determining under what situations these optimality conditions imply anything about the existence of a block structured perturbation matrix Δ satisfying $\det(I - M\Delta) = 0$.

8.1. *Optimality conditions for* $\inf_{D\in\mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$

We want to characterize when $\bar{\sigma}(M) = \inf_{D\in\mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$ that is, when $D := I$ is optimal. Begin with $M \in \mathbf{C}^{n\times n}$, and let its singular value decomposition be

$$M = \sigma_1 UV^* + U_2\Sigma_2V_2^*, \tag{8.1}$$

where $\sigma_1 > 0$ is the maximum singular value of M and has multiplicity r; $U, V \in \mathbf{C}^{n\times r}$; $U_2, V_2 \in \mathbf{C}^{n\times(n-r)}$; $U^*U = V^*V = I_r$; $U_2^*U_2 = V_2^*V_2 = I_{n-r}$; $U^*U_2 = 0$; $V^*V_2 = 0$; and $\Sigma_2 \in \mathbf{R}^{(n-r)\times(n-r)}$ is nonnegative and diagonal with $\sigma_1 I_{n-r} - \Sigma_2 > 0$.

We need some additional notation, in particular

$$\mathbf{Z} := \{D - \tilde{D} : D, \tilde{D} \in \mathbf{D}\}. \tag{8.2}$$

Note that the elements of $\mathbf{Z}$ are not invertible, and in general are of the form (since $d_{S+F} \equiv 1$)

$$\operatorname{diag}[Z_1, \ldots, Z_S, z_{S+1}I_{m_1}, \ldots, z_{S+F-1}I_{m_{F-1}}, 0_{m_F}],$$

where for each $i \le S$, $Z_i = Z_i^* \in \mathbf{C}^{r_i\times r_i}$, and for $j \le F-1$, $z_{S+j} \in \mathbf{R}$. Later, we will use the fact that $\mathbf{Z}$ is a real inner product space, with inner product defined by $P, T \in \mathbf{Z}$

$$\langle P, T\rangle := \sum_{i=1}^{S} \operatorname{trace}(P_iT_i) + \sum_{j=1}^{F-1} p_{S+j}t_{S+j}.$$

For notational purposes, partition U and V

compatibly with $\mathbf{\Delta}$ as

$$U = \begin{bmatrix} A_1 \\ \vdots \\ A_S \\ E_1 \\ \vdots \\ E_F \end{bmatrix}, \quad V = \begin{bmatrix} B_1 \\ \vdots \\ B_S \\ H_1 \\ \vdots \\ H_F \end{bmatrix} \tag{8.3}$$

where $A_i, B_i \in \mathbf{C}^{r_i \times r}$, $E_i, H_i \in \mathbf{C}^{m_i \times r}$. With this notation, and a little manipulation, for any $Z \in \mathbf{Z}$, we can write $\lambda_{\min}(U^*ZU - V^*ZV)$ in terms of inner products in $\mathbf{Z}$,

$$\lambda_{\min}(U^*ZU - V^*ZV) = \min_{\substack{\eta \in \mathbf{C}^r \\ \|\eta\|=1}} \langle Z, P^\eta \rangle, \tag{8.4}$$

where for each $\eta \in \mathbf{C}^n$, $P^\eta \in \mathbf{Z}$ is defined by its block components

$$\begin{aligned} P_i^\eta &:= A_i\eta\eta^*A_i^* - B_i\eta\eta^*B_i^*, \\ p_{S+j}^\eta &:= \eta^*(E_j^*E_j - H_j^*H_j)\eta. \end{aligned} \tag{8.5}$$

Let $\nabla_M \subset \mathbf{Z}$ be the set of all such P^η. That is

$$\nabla_M := \{\operatorname{diag}[P_1^\eta, \ldots, P_S^\eta, p_{S+1}^\eta I_{m_1}, \ldots, p_{S+F-1}^\eta I_{m_{F-1}}, 0_{m_F}] : P_i^\eta, p_{S+j}^\eta \text{ as in (8.5)}, \eta \in \mathbf{C}^r, \|\eta\| = 1\}. \tag{8.6}$$

Although the matrices U and V are not unique, the set ∇_M does not depend on their particular choice. For a given $Z \in \mathbf{Z}$, we have

$$\lambda_{\min}(U^*ZU - V^*ZV) = \min_{P \in \nabla_M} \langle Z, P \rangle. \tag{8.7}$$

Hence, it is the set ∇_M that determines whether or not there is a Z such that

$$\lambda_{\min}(U^*ZU - V^*ZV) > 0.$$

Let the convex hull of a set $\mathcal{V} \subset \mathbf{Z}$ be denoted $\operatorname{co}(\mathcal{V})$.

Theorem 8.1. There exists a $Z \in \mathbf{Z}$ such that $\lambda_{\min}(U^*ZU - V^*ZV) > 0$ if and only if $0 \notin \operatorname{co}(\nabla_M)$.

Proof. This is a consequence of (8.7), and a standard result about convex hulls of sets in inner product spaces (Luenberger, 1969).

Now the optimality condition can be derived. In the proof that follows, note that no appeal to differentiability of eigenvalues is necessary, and all of the steps of the proof are elementary linear algebra. The idea for such a simple approach is from Young (1992) and Poolla (1991).

Theorem 8.2. $\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) = \bar{\sigma}(M)$ if and only if $0 \in \operatorname{co}(\nabla_M)$.

Proof ($\Rightarrow$) Suppose that $0 \notin \operatorname{co}(\nabla_M)$. Choose a matrix $Z \in \mathbf{Z}$ such that

$$\lambda_{\min}(U^*ZU - V^*ZV) > 0.$$

Equivalently,

$$\lambda_{\max}(V^*ZV - U^*ZU) < 0.$$

Now, note that for every $\alpha > 0$

$$\begin{bmatrix} V^* \\ V_2^* \end{bmatrix}[M^*(I - \alpha Z)M - \sigma_1^2(I - \alpha Z)][V \quad V_2],$$

is equal to

$$\begin{bmatrix} \sigma_1^2\alpha(V^*ZV - U^*ZU) & \alpha\sigma_1(\sigma_1 V^*ZV_2 - U^*ZU_2\Sigma_2) \\ \alpha\sigma_1(\sigma_1 V_2^*ZV - \Sigma_2 U_2^*ZU) & \Sigma_2^2 - \sigma_1^2 I + \alpha(\sigma_1^2 V_2^*ZV_2 - \Sigma_2 U_2^*ZU_2\Sigma_2) \end{bmatrix}.$$

Call $T := \sigma_1(\sigma_1 V^*ZV_2 - U^*ZU_2\Sigma_2)$, and $L := (\sigma_1^2 V_2^*ZV_2 - \Sigma_1 U_2^*ZU_2\Sigma_2)$. Using this notation, the matrix in question becomes

$$\begin{bmatrix} \sigma_1^2\alpha(V^*ZV - U^*ZU) & \alpha T \\ \alpha T^* & \Sigma_2^2 - \sigma_1^2 I + \alpha L \end{bmatrix}.$$

Choose $\alpha > 0$ small enough so that the three conditions

$$\begin{gathered} I - \alpha Z > 0, \\ \Sigma_2^2 - \sigma_1^2 I + \alpha L < 0, \end{gathered}$$

$$\sigma_1^2(V^*ZV - U^*ZU) - \alpha T(\Sigma_2^2 - \sigma_1^2 I + \alpha L)^{-1}T^* < 0,$$

are satisfied. This is possible, since $I > 0$, $(V^*ZV - U^*ZU) < 0$, and $\Sigma_2^2 - \sigma_1^2 I < 0$. Using Schur complements, it is clear that for such α, the matrix

$$\begin{bmatrix} \sigma_1^2\alpha(V^*ZV - U^*ZU) & \alpha T \\ \alpha T^* & \Sigma_2^2 - \sigma_1^2 I + \alpha L \end{bmatrix} < 0.$$

This implies that

$$[M^*(I - \alpha Z)M - \sigma_1^2(I - \alpha Z)] < 0.$$

Define $D := I - \alpha Z$, and note that

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) \le \bar{\sigma}(D^{1/2}MD^{-1/2}) < \sigma_1 = \bar{\sigma}(M),$$

as desired.

($\Leftarrow$) Suppose that $\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) < \bar{\sigma}(M)$. Choose $D \in \mathbf{D}$ such that $\bar{\sigma}(D^{1/2}MD^{-1/2}) < \bar{\sigma}(M)$. Define $Z \in \mathbf{Z}$ via $Z := I - D$. Note that

$$M^*DM - \sigma_1^2 D = M^*(I - Z)M - \sigma_1^2(I - Z) < 0.$$

Hence, for all $\eta \in \mathbf{C}^r$, with $\|\eta\| = 1$, we have

$$\begin{aligned} 0 &> \eta^*V^*[M^*(I - Z)M - \sigma_1^2(I - Z)]V\eta \\ &= \sigma_1^2[\eta^*U^*(I - Z)U\eta - \eta^*V^*(I - Z)V\eta] \\ &= \sigma_1^2\eta^*(V^*ZV - U^*ZU)\eta. \end{aligned}$$

Hence, this $Z := I - D \in \mathbf{Z}$ satisfies

$$\lambda_{\max}(V^*ZV - U^*ZU) < 0,$$

which is equivalent to

$$\lambda_{\min}(U^*ZU - V^*ZV) > 0.$$

By Theorem 8.1, it must be that $0 \notin \mathrm{co}\,(\nabla_M)$, as desired.

8.2. *Connecting* μ *with* $\bar{\sigma}$(M)

The convex hull of ∇_M determines whether or not $D := I$ is the optimum scaling. Following Doyle (1982) we ask, "what is true about M if $0 \in \nabla_M$?" Since $\nabla_M \subset \mathrm{co}\,(\nabla_M)$ certainly $D := I$ is optimal, but is anything else true? The answer links the upper bound and μ.

Theorem 8.3. Let $M \in \mathbf{C}^{n \times n}$ be given, along with a block structure $\mathbf{\Delta}$, and define ∇_M accordingly (equations (8.1), (8.3), (8.5) and (8.6)). Then, $\bar{\sigma}(M) = \mu(M)$ if and only if $0 \in \nabla_M$.

Proof. The following four statements are equivalent.

(1) $0 \in \nabla_M$.
(2) There exists $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$ and $Q \in \mathbf{Q}$ with $QU\eta = V\eta$.
(3) There exists $\xi \in \mathbf{C}^n$, $\|\xi\| = 1$ and $Q \in \mathbf{Q}$ with $QM\xi = \bar{\sigma}\xi$.
(4) $\bar{\sigma}(M) = \mu_{\Delta}(M)$.

$1 \to 2$: From the definition of ∇_M, (8.6), $0 \in \nabla_M$ implies that for some $\eta \in \mathbf{C}^r$, $\|\eta\| = 1$,

$$\begin{aligned} A_i\eta\eta^*A_i^* - B_i\eta\eta^*B_i^* &= 0, \quad i \le S, \\ \eta^*(E_j^*E_j - H_j^*H_j)\eta &= 0, \quad j \le F - 1. \end{aligned} \tag{8.8}$$

Obviously, for $i \le S$, there is a phase $e^{j\theta_i}$ such that $e^{j\theta_i}A_i\eta = B_i\eta$. For $j \le F - 1$, $\|E_j\eta\| = \|H_j\eta\|$, so there exists a unitary matrix Q_j such that $Q_jE_j\eta = H_j\eta$. The only thing left is the last full block. Since $\|U\eta\| = \|V\eta\|$ we must have $\|E_F\eta\| = \|H_F\eta\|$. This gives a unitary matrix Q_F with $Q_FE_F\eta = H_F\eta$. Arranging the phases and Qs in a block diagonal fashion gives Statement (2).

$2 \to 1$: This follows along the lines of $1 \to 2$.

$2 \to 3$: The matrix M has a SVD of $M = \bar{\sigma}UV^* + U_2\Sigma_2V_2^*$. Hence $QM(V\eta) = \bar{\sigma}QU\eta = \bar{\sigma}V\eta$. Defining $\xi = V\eta$ gives Statement (3).

$3 \to 2$: A SVD of QM is $QM = \bar{\sigma}(QU)V^* + (QU_2)\Sigma_2V_2^*$. If $QM\xi = \bar{\sigma}\xi$, then ξ must lie in the subspace spanned by the right singular vectors associated with $\bar{\sigma}$. Hence there is a vector η satisfying $\xi = V\eta$. Obviously $\|\eta\| = 1$ and

$$QU\eta = QUV^*\xi = \frac{1}{\bar{\sigma}}QM\xi = \xi = V\eta. \tag{8.9}$$

$3 \to 4$: $QM\xi = \bar{\sigma}\xi$ implies that $\mu_{\Delta}(M) = \max_{Q \in \mathbf{Q}} \rho(QM) \ge \rho(QM) \ge \bar{\sigma}(M)$. However, $\bar{\sigma}$ is always an upper bound for μ, hence we must have equality.

$4 \to 3$: This is clear, since $\max_{Q \in \mathbf{Q}} \rho(QM) = \mu_{\Delta}(M)$.

Theorem 8.3 can be used to relate the upper bound and $\mu_{\Delta}(M)$. In particular, we consider block structure $\mathbf{\Delta}$ that have the following property: for all $W \in \mathbf{C}^{n \times n}$, $0 \in \mathrm{co}\,(\nabla_W)$ always implies $0 \in \nabla_W$. Note that while this property is stated in terms of ∇_W, is is actually a property of the underlying block structure. We will say that a block structure satisfying this property is μ-simple. In Section 9, we will completely characterize which block structures are μ-simple, and which block structures are not. For now, we prove that μ-simple block structures always have μ equal to the upper bound.

Theorem 8.4. Suppose the block structure $\mathbf{\Delta}$ is μ-simple. Then, for every $M \in \mathbf{C}^{n \times n}$,

$$\mu_{\Delta}(M) = \inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}).$$

Proof. Let $\beta = \inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$. Let D_k be a sequence in $\mathbf{D}$ such that $\bar{\sigma}(D_k^{1/2}MD_k^{-1/2})$ converges to β as $k \to \infty$. Denote $W_k = D_k^{1/2}MD_k^{-1/2}$. Since the sequence W_k is bounded, it has a convergent subsequence with limit W. Obviously, by continuity of $\bar{\sigma}$ and μ, $\bar{\sigma}(W) = \beta$ and $\mu_{\Delta}(M) = \mu_{\Delta}(W)$. We claim that $0 \in \mathrm{co}\,(\nabla_W)$. If not, then there exist $D \in \mathbf{D}$ and $\epsilon > 0$ such that $\bar{\sigma}(D^{1/2}WD^{-1/2}) = \beta - \epsilon$. Choose k so that $\|W_k - W\| < \epsilon/2\sqrt{\kappa(D)}$, where $\kappa(\cdot)$ denotes condition number. Then

$$\|D^{1/2}(W_k - W)D^{-1/2}\| < \frac{\epsilon}{2},$$

which yields

$$\|D^{1/2}W_kD^{-1/2}\| < \beta - \frac{\epsilon}{2}.$$

This contradicts that β was the infimum, thus indeed $0 \in \mathrm{co}\,(\nabla_W)$. By hypothesis, this means $0 \in \nabla_W$ so by Theorem 8.3, $\mu_{\Delta}(W) = \bar{\sigma}(W)$. Recalling continuity, we get $\mu_{\Delta}(M) = \beta$ as desired.

Consider the minimization over the Ds. Since we are minimizing the maximum singular value, the top singular values tend to coalesce, so that at the minimum, the multiplicity of $\bar{\sigma}$ is greater than or equal to two. This is typical of any

"min–max" problem. Suppose though, that at the minimum, $\bar{\sigma}(D_{\mathrm{opt}}^{1/2}MD_{\mathrm{opt}}^{-1/2})$ was distinct. Obviously, since we are at a minimum, we must have $0 \in \mathrm{co}(\nabla)$. But if the multiplicity of $\bar{\sigma}$ is only one, then ∇ is a single point, hence $\nabla = \{0\}$. This reasoning gives:

Corollary 8.5. If, at the minimum of $\bar{\sigma}(D^{1/2}MD^{-1/2})$ the maximum singular value has multiplicity of one, then $\mu(M) = \min_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$.

9. PROPERTIES OF ∇

In this section, we study the convexity properties of the set ∇, since the relationship between $\mu_\Delta(M)$ and $\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$ depend on the relationship between ∇ and co (∇). It will be shown, that for some block structures Δ, the implication

$$0 \in \mathrm{co}\,(\nabla_W) \rightarrow 0 \in \nabla_W,$$

holds for every complex matrix W of appropriate dimensions. Hence, for those block structures,

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) = \mu_\Delta(M),$$

for every matrix M. Likewise, for other block structures, specific matrices can be constructed for which the upper bound can be shown to be greater than μ. The upper bound may be equal to μ for certain matrices (see Theorem 8.3 for example) but in general, the upper bound is not equal to μ.

These upcoming results are summarized in the Table 1, which indicates section numbers for the accompanying derivation or example. Note that the $(S = 0,\ F = 1)$ entry is trivial, and the $(S = 1,\ F = 0)$ entry implies that for any $M \in \mathbf{C}^{n \times n}$

$$\rho(M) = \inf_{\substack{D \in \mathbf{C}^{n \times n} \\ D - D^* > 0}} \bar{\sigma}(D^{1/2}MD^{-1/2}),$$

which is a well-known fact.

Before beginning, we make a notational change, for ease of exposition. Although the set ∇ was defined as a subset of block diagonal, $n \times n$ Hermitian matrices, in this section we identify ∇ with the set $\mathbf{H}^{r_1} \times \mathbf{H}^{r_2} \times \cdots \times \mathbf{H}^{r_S} \times \mathbf{R}^{F-1}$.

9.1. S = 0, F = 2

The situation with two full blocks is relatively simple. Referring back to (8.5), ∇ will always have the form

$$\nabla = \{\eta^*(E^*E - F^*F)\eta : \eta \in \mathbf{C}^r,\ \|\eta\| = 1\}, \qquad (9.1)$$

for some given $r > 0$ and $E, F \in \mathbf{C}^{m_i \times r}$. Since $E^*E - F^*F$ is Hermitian, ∇ is just a closed interval in the real line. Obviously, this is always convex, so if $0 \in \mathrm{co}\,(\nabla)$, in fact, $0 \in \nabla$. Hence by Theorem 8.4:

Theorem 9.1. If Δ consists of two full blocks ($S = 0$, $F = 2$), then

$$\mu_\Delta(M) = \inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})$$

Remark 9.2. The two block case was first solved by Redheffer (1959) with a quite different approach involving the use of Schauder's fixed point theorem (Dunford and Schwartz, 1958, 1963).

9.2. S = 0, F = 4 (*Morton and Doyle*, 1985)

Consider the case when Δ consists of four 1×1 blocks, so $S = 0$, $F = 4$, and $m_j = 1$ for each j. Let a, b, and c be positive real numbers, d and f be complex numbers, and ψ_1 and ψ_2 be real numbers. Define matrices $U, V \in \mathbf{C}^{4 \times 2}$ by

$$U = \begin{bmatrix} 0 \\ b \\ jc \\ f \end{bmatrix} \qquad V = \begin{bmatrix} 0 & a \\ b & -b \\ c & -jc \\ e^{j\psi_1}f & e^{j\psi_2}d \end{bmatrix}$$

For the time being, suppose that these are both unitary matrices, so that $U^*U = V^*V = I_2$. Later we will actually assign the correct values, but at the moment we just assume this is already done. Then define $M \in \mathbf{C}^{4 \times 4}$ by

$$M := UV^*. \qquad (9.2)$$

With the assumptions of unitariness on U and V,

TABLE 1. GUARANTEED EQUALITY BETWEEN μ AND THE UPPER BOUND

	F=0	F=1	F=2	F=3	F=4
S=0		YES	YES Section 9.1	YES Section 9.3	NO Section 9.2
S=1	YES	YES Section 9.4	NO Section 9.6	NO	NO
S=2	NO Section 9.5	NO	NO	NO	NO

(9.2) is a singular value decompostion of M. M has two singular values at 1, and two singular values at 0. With respect to the block structure $\boldsymbol{\Delta}$ that we have defined, what properties does the set ∇_M have? In particular:

- Is $0 \in \mathrm{co}(\nabla_M)$? If so, then
$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) = 1,$$
otherwise, it is less than one.
- Is $0 \in \nabla_M$? If so, then $\mu(M) = \bar{\sigma}(M) = 1$, otherwise it is less than one.

Since the multiplicity of the maximum singular value is two, we can parametrize all unit vectors in $\mathbf{C}^2$, and get a parametric representation of ∇_M. It is easy to see that any vector $\eta \in \mathbf{C}^2$, with $\|\eta\| = 1$ is of the form

$$\eta = \begin{bmatrix} e^{j\phi_1} \cos\theta \\ e^{j\phi_2} \sin\theta \end{bmatrix},$$

for some real ϕ_1, ϕ_2, and θ. As it turns out, ∇_M depends only on the difference $\phi_1 - \phi_2$, which we will denote as ϕ. Simply plugging in for the definition of ∇_M from Section 8, we get

$$\nabla_M = \left\{ \begin{bmatrix} a^2(\cos^2\theta - \sin^2\theta) \\ 4b^2 \sin\theta\cos\theta\cos\phi \\ 4c^2 \sin\theta\cos\theta\sin\phi \end{bmatrix} \in \mathbf{R}^3 : \phi, \theta \in \mathbf{R} \right\} \subset \mathbf{R}^3. \quad (9.3)$$

It is apparent that $0 \neq \nabla_M$. That would require (from the first coordinate in (9.3)) that $\theta = [(2n+1)/4]\pi$, for some integer n. The second and third coordinates being zero would then require both $\cos\phi = 0$ and $\sin\phi = 0$, which is impossible. Hence $0 \notin \nabla_M$, and $\mu(M) < 1$.

On the other hand, setting $\theta = 0$, and then $\theta = \pi/2$, gives that both $[a^2\ 0\ 0]^T$ and $[-a^2\ 0\ 0]^T$ are elements of ∇_M. Consequently, $0 \in \mathrm{co}(\nabla_M)$. Therefore

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) = \bar{\sigma}(M) = 1.$$

In order to complete the counter-example, we must choose the free variables so that U and V in (9.2) are unitary, as we said we could.

Set $\gamma = 3 + \sqrt{3}$ and $\beta = \sqrt{3} - 1$ and define $a = \sqrt{2/\gamma}$, $b = 1/\sqrt{\gamma}$, $c = 1/\sqrt{\gamma}$, $d = -\sqrt{\beta/\gamma}$, $f = (1+j)\sqrt{1/\gamma\beta}$, $\psi_1 = -\pi/2$ and $\psi_2 = \pi$. Some algebra later, we conclude that ∇_M is the set of all $x \in \mathbf{R}^3$, such that $\|x\| = 2/(3+\sqrt{3})$. Obviously, $0 \notin \nabla_M$, but $0 \in \mathrm{co}(\nabla_M)$. Extensive searching over the set $\mathbf{Q}$ in the lower bound formula has revealed that for M defined above, $\mu_{\boldsymbol{\Delta}}(M)$ is approximately 0.87326. This counter-example proves that for every block structure $\boldsymbol{\Delta}$ satisfying $S + F \geq 4$, there exist matrices M with

$$\mu_{\boldsymbol{\Delta}}(M) < \inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}).$$

9.3. S = 0, F = 3 (*Doyle*, 1982)

In this problem, for every matrix M, $\nabla_M \subset \mathbf{R}^2$, of the form

$$\left\{ \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \end{bmatrix} \in \mathbf{R}^2 : \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\} \subset \mathbf{R}^2, \quad (9.4)$$

for some integer r, and Hermitian matrices H_1 and $H_2 \in \mathbf{C}^{r \times r}$. In Doyle (1982), it is shown that this set is always convex, so that the upper bound is exactly equal to $\mu_{\boldsymbol{\Delta}}(M)$. For completeness, the results needed to prove this are stated below.

Begin with some notation from Doyle (1982). For any positive integer r, define the sets $P^r := \{x \in \mathbf{C}^r : \|x\| = 1\}$ and $\mathbf{S}^r := \{v \in \mathbf{R}^{r+1} : \|v\| = 1\}$. If $H_1, H_2, \ldots, H_q$ are Hermitian matrices in $\mathbf{C}^{r \times r}$, define a function $f_H : P^r \to \mathbf{R}^q$ by

$$f_H(\eta) := \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \\ \vdots \\ \eta^* H_q \eta \end{bmatrix} \in \mathbf{R}^q, \quad (9.5)$$

for each $\eta \in P^r$.

Lemma 9.3. Let q be a positive integer. Let $a_i, c_i \in \mathbf{R}$, and $b_i \in \mathbf{C}$ for $i = 1, \ldots, q$. For each i, define a Hermitian 2×2 matrix H_i by

$$H_i := \begin{bmatrix} a_i & b_i \\ \bar{b}_i & c_i \end{bmatrix}.$$

Then there exists a vector $d \in \mathbf{R}^q$ and a matrix $V \in \mathbf{R}^{q \times 3}$ such that

$$f_H(P^2) := \{f_H(\eta) : \eta \in P^2\} = \{d + Vu : u \in \mathbf{S}^2\},$$

where f_H is defined in (9.5).

Lemma 9.4. Let $d \in \mathbf{R}^2$ and $V \in \mathbf{R}^{2 \times 3}$. Then the set $\{d + Vu : u \in \mathbf{S}^2\} \subset \mathbf{R}^2$ is convex.

Hence, for $q = 2$ and $r = 2$, the set $f(P^2) \in \mathbf{R}^2$ is convex. For a block structure with $S = 0$, $F = 3$, the set ∇ is always of the form $f(P^r) \subseteq \mathbf{R}^2$ (i.e. $q = 2$). Recall though, that r is the multiplicity of the maximum singular value. Conceivably, this can be any positive number, hence the above reasoning needs to be generalized for $r > 2$.

Theorem 9.5. Let r be any positive integer. Let $H_1, H_2 \in \mathbf{C}^{r \times r}$ be Hermitian matrices. Then the

set

$$f_H(P^r) = \left\{ \begin{bmatrix} \eta^* H_1 \eta \\ \eta^* H_2 \eta \end{bmatrix} \in \mathbf{R}^2 : \eta \in \mathbf{C}^r, \|\eta\| = 1 \right\}, \quad (9.6)$$

is convex.

9.4. S = 1, F = 1

Consider a block structure of one repeated scalar block, and one full block, $S = F = 1$. Recall the definition of ∇_M, equation (8.6). With this structure, the set ∇_M will always be of the form

$$\nabla = \{A\eta\eta^*A^* - B\eta\eta^*B^* : \eta \in \mathbf{C}^r, \|\eta\| = 1\}, \quad (9.7)$$

for some given $r > 0$ and $A, B \in \mathbf{C}^{n \times r}$. It is easy to see that in general, ∇ is not convex. For instance, take $A = I$ and $B = 0$. However the following is always true.

Theorem 9.6. Let ∇ be defined as in (9.7), for arbitrary matrices A and B of appropriate dimensions. If $0 \in \mathrm{co}(\nabla)$, then $0 \in \nabla$.

Proof. Suppose that $0 \in \mathrm{co}(\nabla)$. Then, for some integer p, there exist nonnegative α_i with $\sum_{i=1}^p \alpha_i = 1$ and vectors $\eta_i \in \mathbf{C}^r$ with $\|\eta_i\| = 1$ such that

$$\sum_{i=1}^{p} \alpha_i (A\eta_i\eta_i^*A^* - B\eta_i\eta_i^*B^*) = 0, \quad (9.8)$$

which is rewritten as

$$A\left(\sum_{i=1}^{p} \alpha_i\eta_i\eta_i^*\right)A^* = B\left(\sum_{i=1}^{p} \alpha_i\eta_i\eta_i^*\right)B^*. \quad (9.9)$$

Since the α_i are nonnegative, and not all 0, the dyad summation in (9.9) is a positive semidefinite matrix that is not zero. Let $X^{1/2}$ be its Hermitian, positive semidefinite square root. Therefore $AX^{1/2}X^{1/2}A^* = BX^{1/2}X^{1/2}B^*$. Hence, there is a unitary matrix V such that $AX^{1/2} = BX^{1/2}V$. Let v be an eigenvector of V (with eigenvalue $e^{j\theta}$) such that $X^{1/2}v \neq 0$, and define $u := X^{1/2}v$. Note that u is nonzero. This gives $Au = e^{j\theta}Bu$, which implies that $0 \in \nabla$.

This theorem, along with the LFT machinery developed earlier, can be used to give μ-based derivations of several standard results in linear systems theory, such as the Bounded Real Lemma and the Kalman–Yacobovitch–Popov Lemma. These are relatively straightforward exercises and will not be pursued further here.

9.5. S = 2, F = 0

The block structure considered has $S = 2$ and $F = 0$. A cumbersome example which established the same conclusion appeared in Anderson *et al.* (1986). The example presented here is minimal, in the sense that no smaller problem ("smaller" meaning dimension of blocks and M) could be a counter-example (by results in Section 9.4). The dimension of each repeated scalar block is two.

(a) Let $a \in (0, 1)$ and $\gamma \in (0, 1)$ be given. Define the matrix $M \in \mathbf{R}^{4\times4}$ by

$$M := \begin{bmatrix} 0 & 1 & 0 & 1 \\ \gamma & 0 & \gamma & 0 \\ 2a & 0 & a & 0 \\ 0 & -2a & 0 & -a \end{bmatrix}.$$

Define a block structure $\mathbf{\Delta}_2 := \{\delta_2 I_2 : \delta_2 \in \mathbf{C}\}$.

(b) For all $\Delta_2 \in \mathbf{B}_2$ the LFT $\mathcal{S}(M, \Delta_2)$ is well defined, and appears as

$$\mathcal{S}(M, \Delta_2) = \begin{bmatrix} 0 & \dfrac{1 - a\delta_2}{1 + a\delta_2} \\ \gamma\dfrac{1 + a\delta_2}{1 - a\delta_2} & 0 \end{bmatrix} \quad (9.10)$$

Note that for each such $\Delta_2 = \delta_2 I_2$, the structural radius of $\mathcal{S}(M, \Delta_2)$ is simply $\sqrt{\gamma}$, which by assumption is less than one. With respect to the structure

$$\mathbf{\Delta} := \{\mathrm{diag}[\delta_1 I_2, \delta_2 I_2] : \delta_i \in \mathbf{C}\},$$

Theorem 4.3 implies that $\mu_{\mathbf{\Delta}}(M) < 1$.

(c) Consider the product of two linear fractional transformations with different Δ_2 in $\mathbf{B}_2$.

$$\mathcal{S}(M, -I_2)\mathcal{S}(M, I_2) = \begin{bmatrix} \gamma\dfrac{(1+a)^2}{(1-a)^2} & 0 \\ 0 & \gamma\dfrac{(1-a)^2}{(1+a)^2} \end{bmatrix}$$

For any $\gamma \in (0, 1)$, it is easy to choose $a \in (0, 1)$ so that the spectral radius of the above product is greater than one. For such choices, then, we must have

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) \geq 1,$$

where $\mathbf{D}$ be the scaling set associated with $\mathbf{\Delta}$. Otherwise, by Lemma 4.9, the spectral radius of any product of these LFTs would be less than one.

Remark 9.7. A deeper analysis can show that by proper choice of γ and a, the value of

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}),$$

can be made arbitrarily close to $1 + \sqrt{2}$ while $\mu_{\mathbf{\Delta}}(M) < 1$.

9.6. $S=1, F=2$

Next is an example for a block structure with $S=1$ and $F=2$. Again, the example here is minimal—no smaller could be a counter-example for this block structure. It is broken down into eight facts.

(a) Let $\mathbf{\Delta}_2 = \{\text{diag}[\delta_1, \delta_2] : \delta_i \in \mathbf{C}\}$. Then for any complex $\tau \neq 0$,

$$\mu_{\Delta_2}\left(\begin{bmatrix} 0 & \frac{1}{\tau} \\ \tau & 0 \end{bmatrix}\right) = 1.$$

(b) Let $a \in \mathbf{C}$ with $|a| < 1$. Define G on $|\delta| \le 1$ as

$$G(\delta) := \begin{bmatrix} 0 & \dfrac{1+a\bar{\delta}}{1-a\delta} \\ \dfrac{1-a\delta}{1+a\delta} & 0 \end{bmatrix} \tag{9.11}$$

Note that everywhere in the unit disk, G is defined and looks like $\begin{bmatrix} 0 & 1/\tau \\ \tau & 0 \end{bmatrix}$. Hence from (a)

$$\sup_{|\delta|\le 1} \mu_{\Delta_2}(G(\delta)) = 1.$$

(c) $G(\delta)$ in (9.11) can be written as a linear fractional transformation. In particular, define the matrix M by

$$M := \begin{bmatrix} -a & 0 & -2a & 0 \\ 0 & a & 0 & 2a \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \tag{9.12}$$

It is simple to verify for each $|\delta| \le 1$, $G(\delta) = \mathcal{S}(\delta I_2, M)$.

(d) Define $\mathbf{\Delta}_1 := \{\delta I_2 : \delta \in \mathbf{C}\}$, and $\mathbf{\Delta}_2 := \{\text{diag}[\delta_1, \delta_2] : \delta_i \in \mathbf{C}\}$ and $\mathbf{\Delta}$ the augmentation of the two sets. Certainly $\mu_{\mathbf{\Delta}}(M)$ makes sense (dimensions are compatible), and $\mu_{\mathbf{\Delta}}(M) \ge 1$, since $\mu_2(M_{22}) = 1$. Using (b) and (c), and Theorem 4.3, gives $\mu_{\mathbf{\Delta}}(M) \le 1$. Therefore $\mu_{\mathbf{\Delta}}(M) = 1$.

(e) Define the usual scaling sets $\mathbf{D}_1$ and $\mathbf{D}_2$ compatible with $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. For any $D_2 \in \mathbf{D}_2$

$$D_2^{1/2}\mathcal{S}(\delta I_2, M)D_2^{-1/2}$$
$$\begin{bmatrix} 0 & \sqrt{\dfrac{d_1}{d_2}\dfrac{1+a\delta}{1-a\delta}} \\ \sqrt{\dfrac{d_2}{d_1}\dfrac{1-a\delta}{1+a\delta}} & 0 \end{bmatrix} \tag{9.13}$$

Hence, with some simple calculus, it is easy to verify that for any $\beta \ge 1$,

$$\sup_{|\delta|\le 1/\beta} \bar{\sigma}(D_2^{1/2}\mathcal{S}(\delta I_2, M)D_2^{-1/2}) \ge \frac{\beta + |a|}{\beta - |a|}.$$

(f) *Fact*. Let $\gamma > 0$. If there is a $\Delta_1 \in \mathbf{\Delta}_1$, $\bar{\sigma}(\Delta_1) \le 1/\gamma$ such that

- $I - M_{11}\Delta_1$ is invertible;
- $\bar{\sigma}[\mathcal{S}(\Delta_1, M)] \ge \gamma$;

then

$$\inf_{D_1 \in \mathbf{D}_1} \bar{\sigma}\left[\begin{pmatrix} D_1^{1/2} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \times \begin{pmatrix} D_1^{-1/2} & 0 \\ 0 & I \end{pmatrix}\right] \ge \gamma.$$

This fact is simply the contra-positive of Lemma 4.9.

(g) If we choose a $\beta \ge 1$ such that

$$\beta + |a|/\beta - |a| \ge \beta,$$

then we can apply the results from (e) and (f) above to conclude that

$$\inf_{D_1 \in \mathbf{D}_1} \bar{\sigma}\left[\begin{pmatrix} D_1^{1/2} & 0 \\ 0 & D_2^{1/2} \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \times \begin{pmatrix} D_1^{-1/2} & 0 \\ 0 & D_2^{-1/2} \end{pmatrix}\right] \ge \beta.$$

This logic is as follows: first suppose β is chosen so that $(\beta + |a|)/(\beta - |a|) \ge \beta$. Then from equation (9.13) we know that for every $D_2 \in \mathbf{D}_2$, there is a $\delta \in \mathbf{C}$ with $|\delta| \le 1/\beta$ such that

$$\bar{\sigma}(D_2^{1/2}\mathcal{S}(\delta I_2, M)D_2^{-1/2}) \ge \beta.$$

This satisfies the conditions of (f), therefore, for each $D_2 \in \mathbf{D}_2$

$$\inf_{D_1 \in \mathbf{D}_1} \bar{\sigma}\left[\begin{pmatrix} D_1^{1/2} & 0 \\ 0 & D_2^{1/2} \end{pmatrix}\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \times \begin{pmatrix} D_1^{-1/2} & 0 \\ 0 & D_2^{-1/2} \end{pmatrix}\right] \ge \beta. \tag{9.14}$$

Carrying out the infimum over $\mathbf{D}_2$ in (9.14) yields

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2}) \ge \beta,$$

where $\mathbf{D}$ is the diagonal augmentation of $\mathbf{D}_1$ and $\mathbf{D}_2$. Therefore the question becomes: "What is the largest β such that

$$\beta + |a|/\beta - |a| \ge \beta?".$$

Simple algebra gives the largest β as

$$\beta = \overline{|a| + 1 + \sqrt{|a|^2 + 6|a| + 1}}$$

Note that as $|a| \nearrow 1$, the quantity $\beta \nearrow 1 + \sqrt{2}$.

(h) In summary let $\epsilon > 0$. Choose $a \in \mathbf{C}$, $|a| < 1$ such that

$$\overline{|a| + 1 + \sqrt{|a|^2 + 6|a| + 1}} > 1 + \sqrt{2} - \epsilon.$$

Define M as in (9.12). Then, with respect to

the augmented structure described in (d), $\mu_{\Delta}(M)=1$ but $\inf_{D\in\mathbf{D}} \bar{\sigma}(D^{1/2}MD^{-1/2})>1+\sqrt{2}-\epsilon$.

9.7. $\mathbf{M}\in\mathbf{R}^{n\times n}$, $\mathrm{S}=0$, $\mathrm{F}=2$

If M is real, and the block structure $\mathbf{\Delta}$ consists of two full blocks, then the smallest perturbation $\Delta\in\mathbf{\Delta}$ making $I-M\Delta$ singular will actually be a real matrix, rather than complex. The proof is rather simple using the ∇ set. We also note that this result can be found in Redheffer (1959). An elementary result from linear algebra is the key idea.

Lemma 9.8. Suppose r is a positive integer, and $H\in\mathbf{R}^{r\times r}$ is symmetric. Then

$$\{\eta^*H\eta:\eta\in\mathbf{C}^r,\ \|\eta\|=1\}$$
$$=\{\eta^TH\eta:\eta\in\mathbf{R}^r,\ \|\eta\|=1\}.$$

In view of this, suppose $M\in\mathbf{R}^{(n+m)\times(n+m)}$, $\mathbf{\Delta}=\{\mathrm{diag}\,[\Delta_1,\Delta_2]:\Delta_1\in\mathbf{C}^{n\times n},\ \Delta_2\in\mathbf{C}^{m\times m}\}$ and the optimal D scaling has been computed. Note that in a two-block problem, if the infimum is not achieved, then it must be that either $M_{12}=0$ or $M_{21}=0$, and $\mu_{\Delta}(M)=\max\{\bar{\sigma}(M_{11}),\bar{\sigma}(M_{22})\}$. Then, with singular vectors, it is possible to construct a real perturbation of the form $\Delta:=\mathrm{diag}\,[\Delta_1,0]$ or $\Delta:=\mathrm{diag}\,[0,\Delta_2]$ such that $I-M\Delta$ is singular, and $\bar{\sigma}(\Delta)=1/\mu_{\Delta}(M)$. Next, consider the case when the infimum is achieved. Let D be the optimal scaling, and define $W:=D^{1/2}MD^{-1/2}$. Since the optimal D scaling is of the form $\mathrm{diag}\,[d_1I_n, I_m]$, where $d_1>0$, it is clear that W is still real. Hence, $0\in\mathrm{co}\,(\nabla_W)$. Then, for some $U\in\mathbf{R}^{(n-m)\times r}$, $V\in\mathbf{R}^{(n+m)\times r}$,

$$W=\sigma_1UV^*+u_2\Sigma_2V_2^*,$$

and

$$\mu_{\Delta}(M)=\sigma_1=\bar{\sigma}(W).$$

If U and V are partitioned with respect to the block structure as

$$U=\begin{bmatrix}E_1\\E_2\end{bmatrix},\quad V=\begin{bmatrix}F_1\\F_2\end{bmatrix},$$

then ∇_W is

$$\nabla_W:=\{\eta^*E_1^*E_1-F_1^*F_1)\eta:\eta\in\mathbf{C}^r,\ \|\eta\|=1\}.$$

By assumption, the D scaling is optimal, so $0\in\mathrm{co}\,(\nabla_W)=\nabla_W$. Using the lemma, this implies there is a $\eta\in\mathbf{R}^r$, with $|\eta\|=1$ such that $\|E_i\eta\|=\|F_i\eta\|$ for $i=1,2$. It is easy then to construct real, orthogonal matrices Q_1 and Q_2 such that $Q_iE_i\eta=F_i\eta$. Defining $Q:=\mathrm{diag}\,[Q_1,Q_2]$ yields

$$\det\left(I-\left(\frac{1}{\sigma_1}Q\right)M\right)=0,$$

which shows what we had claimed—in the two-block (full blocks) μ problem, with M real, the minimizing perturbation may be taken to be a real matrix.

The next theorem is a mild generalization of these ideas.

Theorem 9.9. Let $\mathbf{\Delta}_R\subseteq\mathbf{R}^{n\times n}$ be a given structure, and define the following four augmented structures:

$$\mathbf{\Delta}_{rrr}=\{\mathrm{diag}\,[\Delta_R,\Delta_1,\Delta_2]:\Delta_R\in\mathbf{\Delta}_R,\ \Delta_1\in\mathbf{R}^{n_1\times n_1},\ \Delta_2\in\mathbf{R}^{n_2\times n_2}\},$$
$$\mathbf{\Delta}_{rrc}=\{\mathrm{diag}\,[\Delta_R,\Delta_1,\Delta_2]:\Delta_R\in\mathbf{\Delta}_R,\ \Delta_1\in\mathbf{R}^{n_1\times n_1},\ \Delta_2\in\mathbf{C}^{n_2\times n_2}\},$$
$$\mathbf{\Delta}_{rcr}=\{\mathrm{diag}\,[\Delta_R,\Delta_1,\Delta_2]:\Delta_R\in\mathbf{\Delta}_R,\ \Delta_1\in\mathbf{C}^{n_1\times n_1},\ \Delta_2\in\mathbf{R}^{n_2\times n_2}\},$$
$$\mathbf{\Delta}_{rcc}=\{\mathrm{diag}\,[\Delta_R,\Delta_1,\Delta_2]:\Delta_R\in\mathbf{\Delta}_R,\ \Delta_1\in\mathbf{C}^{n_1\times n_1},\ \Delta_2\in\mathbf{C}^{n_2\times n_2}\}.$$

Then for $M\in\mathbf{R}^{(n+n_1+n_2)\times(n+n_1+n_2)}$,

$$\mu_{\mathbf{\Delta}_{rrr}}(M)=\mu_{\mathbf{\Delta}_{rrc}}(M)=\mu_{\mathbf{\Delta}_{rcr}}(M)=\mu_{\mathbf{\Delta}_{rcc}}(M).$$

Proof. The proof follows from the previous discussion and Theorem 4.3.

9.8. $\mathbf{M}\in\mathbf{R}^{n\times n}$, $\mathrm{S}=0$, $\mathrm{F}=3$

Unfortunately, the argument used above breaks down in this case, and no longer may the smallest perturbation be assumed real. The following is from Skogestad (1987) and Packard and Doyle (1990). The perturbation set is 3 1×1 blocks. Let $U,V\in\mathbf{R}^{3\times2}$ be

$$U=\begin{bmatrix}1&\beta\\\gamma&\alpha\\\gamma&-\alpha\end{bmatrix},\quad V=\begin{bmatrix}-\beta&0\\\alpha&-\gamma\\\alpha&\gamma\end{bmatrix}$$

where $\alpha,\gamma,\beta\in\mathbf{R}$, and have been chosen so that $U^TU=V^TV=I_2$ (that is easy to do). Define $M:=UV^T\in\mathbf{R}^{3\times3}$. Then $\bar{\sigma}(M)=1$, and $\eta\in\mathbf{C}^2$, $\|\eta\|=1$, parametrized by

$$\eta=\begin{bmatrix}e^{j\psi}\cos\theta\\e^{j\phi}\sin\theta\end{bmatrix},$$

gives

$$\nabla_M=\left\{\begin{bmatrix}-\beta^2\cos2\theta\\(\gamma^2-\alpha^2)\cos2\theta+4\cos(\psi-\phi)\gamma\alpha\cos\theta\sin\theta\end{bmatrix}:\theta,\psi,\phi\in\mathbf{R}\right\}.$$

It is easy to see that $0\in\nabla_M$, by choosing $\theta=(2n+1)\pi/4$ and $\psi-\phi=(2m+1)\pi/2$ for any integers n, m. The only vectors η which lead

to $0 \in \nabla_M$ are

$$\eta = \begin{bmatrix} \pm j e^{j\phi} \frac{1}{\sqrt{2}} \\ \pm e^{j\phi} \frac{1}{\sqrt{2}} \end{bmatrix}$$

which is always complex. Consequently, the only matrices satisfying $\Delta \in \mathbf{\Delta}$, $\bar{\sigma}(\Delta) = 1$, and $I - M\Delta$ singular are complex perturbations.

9.9. $M \in \mathbf{R}^{n \times n}$, $S = 1$, $F = 1$

Again, the smallest perturbations are in general complex. Suppose $G(z)$ is a stable, nth order, SISO transfer function with $\|G\|_\infty = 1$, $|G(1)| < 1$, and $|G(-1)| < 1$. The state space matrix M of this transfer function will have $\mu_\Delta(M) = 1$, but all of the perturbations $\Delta = \mathrm{diag}\,[\delta_1 I_n, \delta_2]$ satisfying $\bar{\sigma}(\Delta) = 1$, and $\det(I - M\Delta) = 0$ will be complex.

9.10. *Optimal scalings for* $\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2} M D^{-1/2})$ *with* $M \in \mathbf{R}^{n \times n}$

If the matrix M is real, then the minimum point in the convex hull of ∇_M is always real, so each block of the optimal $D \in \mathbf{D}$ can be chosen to be real. The proof is very simple.

Theorem 9.10. Let $\mathbf{D}_R$ be the set of real, symmetric members of $\mathbf{D}$. If M is real, then

$$\inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2} M D^{-1/2}) = \inf_{D_R \in \mathbf{D}_R} \bar{\sigma}(D_R^{1/2} M D_R^{-1/2}). \tag{9.15}$$

Proof. Let $D \in \mathbf{D}$ be given, with $D = D_r + jD_i$, and $\bar{\sigma}(D^{1/2} M D^{-1/2}) < \beta$. Note that $D_r = D_r^T > 0$, $D_r \in \mathbf{D}_R$, and $D_i = -D_i^T$. Then

$$M^T(D_r + jD_i)M - \beta^2(D_r + jD_i) < 0. \tag{9.16}$$

Hence, the real part of (9.16) is also symmetric, negative definite so

$$\lambda_{\max}(M^T D_r M - \beta^2 D_r) < 0,$$

which implies that

$$\bar{\sigma}(D_r^{1/2} M D_r^{-1/2}) < \beta,$$

so the infimums are the same.

10. TRANSFER FUNCTIONS, STATE SPACE MATRICES, μ AND ROBUST PERFORMANCE

In this section we begin by establishing some relationships between transfer function matrices and matrices made up of state space realizations. We have already seen one instance of such a connection. In Section 4, it was proven that a finite dimensional linear system is stable, and has $\|\cdot\|_\infty < 1$ if and only if the structured singular value of the state space system matrix is less than one (recall, the block structure consisted of a repeated scalar block, and a full block). We explore this type of manipulation in more detail. With these relationships established, the robust performance properties of an uncertain linear system are investigated, and stated in terms of structured singular value tests. Finally, the connections between the μ-theory and Riccati equations for testing $\mathcal{H}_\infty$ norm bounds are briefly reviewed.

10.1. *Transfer function matrices and state space matrices*

Let $M \in \mathbf{C}^{(n+m)\times(n+m)}$ be given, partitioned as usual, and define the transfer function matrix

$$G(z) := \mathcal{S}\left(\frac{1}{z} I_n, M\right) = M_{22} + M_{21}(zI_n - M_{11})^{-1} M_{12}.$$

Suppose that $\mathbf{\Delta} \subset \mathbf{C}^{m \times}$ is a block structure. Define $\mathbf{\Delta}_P$ as

$$\mathbf{\Delta}_P := \{\mathrm{diag}\,[\delta_1 I_n, \Delta] : \delta_1 \in \mathbf{C}, \Delta \in \mathbf{\Delta}\}.$$

Applying Theorems 4.3 and 6.5, the following statements are equivalent:

(1) $\rho(M_{11}) < 1$ and $\max_{\theta \in [0, 2\pi]} \mu_\Delta(G(e^{j\theta})) < 1$;

(2) $\rho(M_{11}) < 1$ and $\max_{\theta \in [0, 2\pi]} \mu_\Delta(\mathcal{S}(e^{j\theta} I_n, M)) < 1$;

(3) $\rho(M_{11}) < 1$ and $\max_{\substack{\delta_1 \in \mathbf{C} \\ |\delta_1| \le 1}} \mu_\Delta(\mathcal{S}(\delta_1 I_n, M)) < 1$;

(4) $\mu_{\Delta_P}(M) < 1$.

Hence, the peak value of μ of a frequency response is related to a larger μ problem on the state space matrix of the transfer function in question. This generalizes the example in Section 4, where the $\|\cdot\|_\infty$ norm (maximum singular value across frequency) was considered.

Similar results are possible when the upper bound is used instead of μ. Suppose that $\mathbf{D} \subset \mathbf{C}^{m \times m}$ is the scaling set associated with $\mathbf{\Delta}$, as in (3.6). For any $D \in \mathbf{D}$, define

$$M_D := \begin{bmatrix} M_{11} & M_{12} D^{-1/2} \\ D^{1/2} M_{21} & D^{1/2} M_{22} D^{-1/2} \end{bmatrix}.$$

Also, let

$$\mathbf{\Delta}_\sigma := \mathbf{C}^{m \times m},$$

$$\mathbf{\Delta}_N := \{\mathrm{diag}\,[\delta_1 I_n, \Delta_2] : \delta_1 \in \mathbf{C}, \Delta_2 \in \mathbf{C}^{m \times m}\}.$$

Note that $\mu_{\Delta_\sigma}(\cdot)$ is simply the maximum singular value, and that $\mathbf{\Delta}_N$ is μ-simple. Then, the following are equivalent.

(1) $\rho(M_{11}) < 1$ and $\inf_{D \in \mathbf{D}} \|D^{1/2} G D^{-1/2}\|_\infty < 1$.

(2) $\rho(M_{11})<1$ and

$$\inf_{D\in\mathbf{D}} \max_{\substack{\delta\in\mathbf{C}\\ |\delta|\le 1}} \bar{\sigma}[D^{1/2}\mathcal{S}(\delta I_n, M)D^{-1/2}]<1.$$

(3) $\rho(M_{11})<1$, and

$$\inf_{D\in\mathbf{D}} \max_{\substack{\delta\in C\\ |\delta|\le 1}} \mu_{\Delta_n}(\mathcal{S}(\delta I_n, M_D))<1.$$

(4) $\inf_{D\in\mathbf{D}} \mu_{\Delta_N}(M_D)<1.$

(5) $\inf_{\substack{D\in\mathbf{D}\\ X\in\mathbf{C}^{n\times n},\, X=X^*>0}}$

$$\bar{\sigma}\left(\begin{bmatrix} X^{1/2} & 0\\ 0 & D^{1/2}\end{bmatrix} M \begin{bmatrix} X^{-1/2} & 0\\ 0 & D^{-1/2}\end{bmatrix}\right)<1.$$

Also, if $M\in\mathbf{R}^{(n+m)\times(n+m)}$, then the scalings can be chosen to be real, so that the following are equivalent.

(1) $\rho(M_{11})<1$ and $\inf_{D\in\mathbf{D}} \|D^{1/2}GD^{-1/2}\|_\infty<1$,

(2) $\rho(M_{11})<1$ and $\inf_{D\in\mathbf{D}_R} \|D^{1/2}GD^{-1/2}\|_\infty<1$,

(3) $\inf_{\substack{D\in\mathbf{D}_R\\ X\in\mathbf{R}^{n\times n},\, X=X^T>0}}$

$$\bar{\sigma}\left(\begin{bmatrix} X^{1/2} & 0\\ 0 & D^{1/2}\end{bmatrix} M \begin{bmatrix} X^{-1/2} & 0\\ 0 & D^{-1/2}\end{bmatrix}\right)<1.$$

These relationships are very significant. Consider the simple situation where $\mathbf{D}:=\{I_m\}$, in other words, an unscaled transfer function. The equivalences imply that the linear system is stable, and has $\|\cdot\|_\infty$ norm less than one if and only if there is a state-coordinate transformation ($X^{1/2}$) such that the transformed state space matrix

$$\begin{bmatrix} X^{1/2} & 0\\ 0 & I_m\end{bmatrix} M \begin{bmatrix} X^{-1/2} & 0\\ 0 & I_m\end{bmatrix},$$

is a contraction. This is intimately related to the characterization of $\mathcal{H}_\infty$ norms using Riccati equations (Willems, 1971a). This will be discussed further at the end of this section.

10.2. *State space/frequency domain tests for robust performance*

We begin with a matrix $M\in\mathbf{C}^{(n+n_p+m)\times(n\times n_p+m)}$, partitioned as below, relating several variables of a linear system by

$$\begin{bmatrix} x_{k+1}\\ e_k\\ z_k\end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13}\\ M_{21} & M_{22} & M_{23}\\ M_{31} & M_{32} & M_{33}\end{bmatrix}\begin{bmatrix} x_k\\ d_k\\ w_k\end{bmatrix}. \qquad (10.1)$$

The uncertainty is modeled by a feedback loop from z to w through a structured $\Delta\in\mathbf{\Delta}$, where $\mathbf{\Delta}$ is a prescribed $m\times m$ block structure (note that

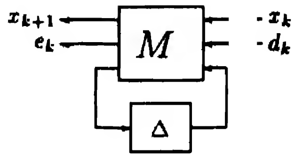


FIG. 10. Uncertain system as an LFT.

we have assumed that the number of disturbance inputs equals the number of errors, and that the perturbation matrices are square—this can all be trivially generalized to include nonsquare situations). Hence, the uncertain system's output error e_k is driven by the input disturbance d_k, and the state equations are given as

$$\begin{bmatrix} x_{k+1}\\ e_k\end{bmatrix} = \mathcal{S}(M,\Delta)\begin{bmatrix} x_k\\ d_k\end{bmatrix}. \qquad (10.2)$$

With respect to the partition, $\mathcal{S}(M,\Delta)$ is

$$\begin{bmatrix} M_{11} & M_{12}\\ M_{21} & M_{22}\end{bmatrix} + \begin{bmatrix} M_{13}\\ M_{23}\end{bmatrix}\Delta(I-M_{33}\Delta)^{-1}[M_{31} \quad M_{32}].$$

This is shown in Fig. 10. Define three augmented block structures, $\mathbf{\Delta}_N$, $\mathbf{\Delta}_S$ and $\mathbf{\Delta}_P$ as

$$\mathbf{\Delta}_N := \{\mathrm{diag}\,[\delta_1 I_n, \Delta_2] : \delta_1\in\mathbf{C}, \Delta_2\in\mathbf{C}^{n_p\times n_p}\},$$
$$\mathbf{\Delta}_S := \{\mathrm{diag}\,[\Delta_N, \Delta] : \Delta_N\in\mathbf{\Delta}_N, \Delta\in\mathbf{\Delta}\},$$
$$\mathbf{\Delta}_P := \{\mathrm{diag}\,[\Delta_2, \Delta] : \Delta_2\in\mathbf{C}^{n_p\times n_p}, \Delta\in\mathbf{\Delta}\},$$

along with the corresponding scaling sets $\mathbf{D}_N$, $\mathbf{D}_S$ and $\mathbf{D}_P$. Motivation for this notation is that the subscript N could mean norm or nominal, S could mean state space, and P could mean performance. We begin with the main result for linear, time-invariant perturbations (Doyle *et al.* 1982; Doyle and Packard, 1987).

Theorem 10.1 (Time-invariant, robust performance). Given the matrices and sets as defined above, the following conditions are equivalent:

(1) there exists a constant $\beta\in[0,1)$ such that for each fixed $\Delta\in\mathbf{B}_\Delta$, the uncertain system (10.2) is well-posed ($I-M_{33}\Delta$ is invertible), stable, and for zero-initial-state-response, the error e satisfies $\|e\|_2\le\beta\,\|d\|_2$;

(2) $\mu_{\Delta_S}(M)<1$ (SSμ test);

(3) $\rho(M_{11})<1$ and $\max_{\theta\in[0,2\pi]} \mu_{\Delta_P}(\mathcal{S}(e^{j\theta}I_n, M))<1$

(FDμ test).

Proof. Introduce two intermediate statements;

1(a) $\mu_\Delta(M_{33})<1$ and $\max_{\Delta\in\mathbf{B}_\Delta} \mu_{\Delta_N}(\mathcal{S}(M,\Delta))<1.$

2(a) $\rho(M_{11})<1$ and $\max_{\delta\in\mathbf{C},|\delta|\le1} \mu_{\Delta_P}(\mathcal{S}(\delta I_n, M))<1.$

The proof that $1 \Leftrightarrow 1(a)$ follows from the definition of stability for a finite dimensional, linear, time-invariant, discrete time system, the relationship between H_∞ norms and l_2 gain, and the equivalence between μ and the $\|\cdot\|_\infty$ norms for transfer functions, as developed in Section 10. Items 1(a), 2 and 2(a) are equivalent by Theorem 4.3, while 2(a) and 3 are equivalent by Theorem 6.5.

Remark 10.2. Item 1 in this theorem is the desired robust performance conclusion. Item 1(a) rephrases Item 1, using the μ characterization of $\|\cdot\|_\infty < 1$. Items 2 and 3 are known, respectively as the "state space μ test" (SSμ) and the "frequency domain μ test" (FDμ). Both of these tests involve computing μ for various matrices. Recall that upper and lower bounds for μ are all that can be computed. Hence, we will investigate the additional conclusions that are possible when the $\bar{\sigma}(D^{1/2}MD^{-1/2})$ upper bound is used to implement the computational tests of Items 2 and 3.

Remark 10.3. The FDμ test is what is most commonly associated with the structured singular value and is often referred to as a μ-plot. It is essentially a Bode magnitude plot with $\mu(\cdot)$ replacing $\bar{\sigma}(\cdot)$ or $|\cdot|$. The SSμ test was introduced in Doyle and Packard (1987).

10.3. *Upper bounds*

Using the $\bar{\sigma}(D^{1/2}MD^{-1/2})$ upper bound in place of μ, we can derive sufficient conditions for robust performance. The resulting state space upper bound test (SSUB) and the frequency domain upper bound test (FDUB) are

$$2' \quad \inf_{D_S \in \mathbf{D}_S} \bar{\sigma}(D_S^{1/2} M D_S^{-1/2}) < 1 \quad \text{(SSUB)},$$

$$3' \quad \max_{\theta \in [0, 2\pi]} \inf_{D_P \in \mathbf{D}_P} \bar{\sigma}[D_P^{1/2} \mathcal{S}(e^{j\theta} I_n, M) D_P^{-1/2}] < 1 \quad \text{(FDUB)}.$$

While the various μ tests given in Theorem 10.1 are all equivalent, these two upper bound tests are very different. In particular, recalling the results from the previous section on scaling a transfer function with a constant similarity transformation, the SSUB condition is actually equivalent to

$$\inf_{D_P \in \mathbf{D}_P} \max_{\theta \in [0, 2\pi]} \bar{\sigma}[D_P^{1/2} \mathcal{S}(e^{j\theta} I_n, M) D_P^{-1/2}] < 1. \tag{10.3}$$

This condition is much stronger than the frequency domain upper bound test, since in (10.3), the same $D_P \in \mathbf{D}_P$ must work for all $\theta \in [0, 2\pi]$. For that reason, we call equation (10.3) the frequency domain constant D test, FDCD. Listed, from strongest to weakest, the various conditions are:

$$\begin{array}{lr}
\inf_{D_S \in \mathbf{D}_S} \bar{\sigma}(D_S^{1/2} M D_S^{-1/2}) < 1 & \text{(SSUB)}, \\
\Updownarrow & \\
\inf_{D_P \in \mathbf{D}_P} \max_{\theta \in [0, 2\pi]} \bar{\sigma}[D_P^{1/2} \mathcal{S}(e^{j\theta} I_n, M) D_P^{-1/2}] < 1 & \text{(FDCD)}, \\
\Downarrow \not\Uparrow & \\
\max_{\theta \in [0, 2\pi]} \inf_{D_P \in \mathbf{D}_P} \bar{\sigma}[D_P^{1/2} \mathcal{S}(e^{j\theta} I_n, M) D_P^{-1/2}] < 1 & \text{(FDUB)}, \\
\Downarrow \not\Uparrow & \\
\max_{\theta \in [0, 2\pi]} \mu_{\Delta_P}(\mathcal{S}(e^{j\theta} I_n, M)) < 1 & \text{(FD}\mu\text{)}, \\
\Updownarrow & \\
\mu_{\Delta_S}(M) < 1 & \text{(SS}\mu\text{)}, \\
\Updownarrow & \\
\text{Condition 1 in Theorem 10.1} &
\end{array}$$

Note that in both instances where the implication is given as $\Downarrow$ rather than $\Updownarrow$, there truly is a gap. Also, there are two such gaps between the state space tests, SSUB and SSμ, while there is only one gap between the frequency domain tests, FDUB and FDμ. The top conditions are the strongest, and are equivalent to a very strong form of robust Lyapunov stability (Boyd and Yang, 1989).

Given that the upper bound is computable, one might ask which test should be used, the state space upper bound test, SSUB (equivalently FDCD), or the frequency domain upper bound test, FDUB? The answer depends on the assumptions that are made about the perturbations. If the SSUB is used and the bound satisfied, then the robust performance conclusion holds for time-varying perturbations (and with proper interpretation, cone bounded nonlinear perturbations).

Theorem 10.4. Let M be given as in (10.1), along with an uncertainty structure $\mathbf{\Delta}$. If there is a $D_S \in \mathbf{D}_S$ such that

$$\bar{\sigma}(D_S^{1/2} M D_S^{-1/2}) = \beta < 1, \tag{10.4}$$

then there exist constants $c_1 \geq c_2 > 0$, such that for all perturbation sequences $\{\Delta_k\}_{k=0}^{\infty}$ with $\Delta_k \in \mathbf{\Delta}$, $\bar{\sigma}(\Delta_k) < 1/\beta$, the time-varying, uncertain system

$$\begin{bmatrix} x_{k+1} \\ e_k \end{bmatrix} = \mathcal{S}(M, \Delta_k) \begin{bmatrix} x_k \\ d_k \end{bmatrix}, \tag{10.5}$$

is zero-input, exponentially stable, and furthermore, if $\{d_k\}_{k=0}^{\infty} \in l_2$, then

$$c_2(1 - \beta^2) \|x\|_2^2 + \|e\|_2^2 \leq \beta^2 \|d\|_2^2 + c_1 \|x_0\|^2.$$

In particular, $\|e\|_2^2 \leq \beta^2 \|d\|_2^2 + c_1 \|x_0\|^2$.

Proof. Note that D_S will appear as $D_S = \operatorname{diag}[D_1, d_2 I, D]$, where $D_1 = D_1^* > 0$, $D_1 \in \mathbf{C}^{n\times n}$. Using equation (10.4), it is easy to show that regardless of $\Delta_k \in \mathbf{\Delta}$, $\bar{\sigma}(\Delta_k) < 1/\beta$, the norms of pertinent vectors satisfy

$$\|D_1^{1/2}x_{k+1}\|^2 + \|e_k\|^2 \le \beta^2(\|D_1^{1/2}x_k\|^2 + \|d_k\|^2).$$

Let c_1 and c_2 be the square roots of the maximum and minimum singular values of D_1. Summing and taking limits yields the final result.

Unfortunately, this test (like the SSμ) does not scale in a convenient manner. In other words, if there is a $D_S \in \mathbf{D}_S$ such that $\bar{\sigma}(D_S^{1/2}MD_S^{-1/2}) = 1.001$, it is impossible to conclude anything about the robust performance characteristics of this system. It is necessary to scale the perturbation channels and/or disturbance channels (this amounts to scaling rows of M to produce a modified system M_{scl}) until a D_S can be found such that $\bar{\sigma}(D_S^{1/2}M_{\text{scl}}D_S^{-1/2}) < 1$, and then robust performance with respect to the scaled down uncertainty and performance norm is guaranteed. For example, let $L = \operatorname{diag}[I_n, 0.8I_{n_p}, (1/1.2)I_m]$. Suppose that there is a $D_S \in \mathbf{D}_S$ such that $\bar{\sigma}(D_S^{1/2}LMD_S^{-1/2}) < 1$. Then it is possible to conclude that for perturbations satisfying $\bar{\sigma}(\Delta_k) \le 0.8$, the error is bounded by, $\|e\|_2^2 \le (1.2)^2\,\|d\|_2^2 + c_1\,\|x_0\|^2$.

Since FDUB is a weaker condition than the SSUB, it is "closer" to the exact condition for robust performance under linear, time-invariant perturbations. Therefore, if the perturbations are better modeled as linear, time-invariant perturbations, this frequency domain test is more appropriate. Also, this test scales, that is, if

$$\max_{\theta\in[0,2\pi]}\ \inf_{D_P\in\mathbf{D}_P} \bar{\sigma}[D_P^{1/2}\mathcal{S}(e^{j\theta}I_n, M)D_P^{-1/2}] = \beta,$$

then the conclusion is that for all $\Delta \in \mathbf{\Delta}$, with $\bar{\sigma}(\Delta) < 1/\beta$, the perturbed system is stable, and the $\|\cdot\|_\infty$ norm of the transfer function from the disturbance to error is $\le\beta$. Hence, peak values other than one still give useful information.

However, if the frequency domain test is used, no general conclusion can be reached about time-varying perturbations (Packard and Doyle, 1990). In Safonov (1984), some connections between the frequency domain test and robust stability to cone bounded nonlinearities are developed.

For reference, continuous-time versions of these theorems, as well as theorems with more sophisticated assumptions about the structured perturbations are found in Chen and Desoer (1982), Doyle *et al.* (1982), Foo and Postlethwaite (1988), Khargonekar and Kaminer (1991), and Bamieh and Dahleh (1992).

10.4. *$\mathcal{H}_\infty$ norms, Riccati equations and LMIs*

We now consider the relationship between the bounds given above and Riccati equations for computing the $\mathcal{H}_\infty$ norm of the discrete time system

$$x_{k+1} = Ax_k + Bu_k,$$
$$y_k = Cx_k + Du_k.$$

Let $M \in \mathbf{C}^{(n+m)\times(n+m)}$ be the block state space matrix of the system

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Assume that A is stable ($\rho(A) < 1$) and define

$$E := A + B(I - D'D)^{-1}D'C,$$
$$G := -B(I - D'D)^{-1}B',$$
$$Q := C'(I - DD')^{-1}C.$$

Suppose E is nonsingular and define a symplectic matrix as

$$S := \begin{bmatrix} E + GE'^{-1}Q & -GE'^{-1} \\ -E'^{-1} & E'^{-1} \end{bmatrix}.$$

It can be shown that the following statements are equivalent:

(a) $\|D + C(zI_n - A)^{-1}B\|_\infty < 1$.

(b) S has no eigenvalues on the unit circle and $\|C(I - A)^{-1}B + D\| < 1$.

(c) $\exists X \ge 0$ with $I - D'D - B'XB > 0$, $(I + GX)^{-1}E$ stable, and

$$E'XE - X - E'XG(I + XG)^{-1}XE + Q = 0.$$

(d) $\exists X > 0$ such that $I - D'D - B'XB > 0$ and

$$E'XE - X - E'XG(1 + XG)^{-1}XE + Q < 0.$$

(e) $\exists X > 0$ such that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}' \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} - \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} < 0.$$

(f) $\exists T$ nonsingular such that

$$\bar{\sigma}\left(\begin{bmatrix} TAT^{-} & TB \\ CT^{-1} & D \end{bmatrix}\right) = \bar{\sigma}\left(\begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} & B \\ & D \end{bmatrix}\begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}^{-1}\right) < 1.$$

Note that (c) is a Riccati equation, and the SSUB in (f) is equal to μ because of the block structure. It is equivalent to (e), which is two LMIs. The connection between (d) and (e) is just the Schur complement formula for positive definite matrices.

11. QUADRATIC LYAPUNOV FUNCTIONS FOR UNCERTAIN SYSTEMS

Some computable results on the quadratic stability of linear systems under structured, linear fractional uncertainty are possible. Again, consider positive integers n and m, and suppose $M \in \mathbf{C}^{(n+m)\times(n+m)}$. Let $\mathbf{\Delta}$ be a structured perturbation set with $\mathbf{\Delta} \subset \mathbf{C}^{m\times m}$. Assume that $\mu_{\mathbf{\Delta}}(M_{22}) < 1$, so that $\mathcal{S}(M, \Delta)$ is well defined for all $\Delta \in \mathbf{B}_{\mathbf{\Delta}}$.

Let $\{\Delta_k\}_{k=0}^{\infty}$ with $\Delta_k \in \mathbf{B}_{\mathbf{\Delta}}$ be given, along with an initial condition $x_0 \in \mathbf{C}^n$, define $x_k \in \mathbf{C}^n$ by the uncertain difference equation

$$x_{k+1} = \mathcal{S}(M, \Delta_k)x_k. \tag{11.1}$$

In this formulation, the matrix M_{11} may be thought of as a nominal state space model and $\Delta_k \in \mathbf{B}_{\mathbf{\Delta}}$ as a norm bounded perturbation from an allowable perturbation class, $\mathbf{\Delta}$. The matrices M_{12}, M_{21}, and M_{22} reflect prior knowledge on how the unknown perturbation affects the nominal dynamics, M_{11}.

Definition 11.1. The pair $(M, \mathbf{\Delta})$ is quadratically stable if there exists a $P \in \mathbf{C}^{n\times n}$, with $P = P^* > 0$, such that

$$\max_{\Delta \in \mathbf{B}_{\mathbf{\Delta}}} \lambda_{\max}([\mathcal{S}(M, \Delta)]^* P \mathcal{S}(M, \Delta) - P) < 0.$$

The definition simply implies that there is a single quadratic Lyapunov function, $V(x) := x^*Px$, that establishes the stability of the entire set

$$\{\mathcal{S}(M, \Delta) : \Delta \in \mathbf{B}_{\mathbf{\Delta}}\}.$$

Equivalently, the definition implies that there is a positive definite $P \in \mathbf{C}^{n\times n}$ such that

$$\max_{\Delta \in \mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}(P^{1/2}\mathcal{S}(M, \Delta)P^{-1/2}) = \gamma < 1.$$

Hence, with respect to a single coordinate change defined by $P^{1/2}$, $\mathcal{S}(M, \Delta_k)$ is always a contraction, regardless of $\Delta_k \in \mathbf{B}_{\mathbf{\Delta}}$. As the uncertain system in (11.1) evolves, the Euclidean norm of $P^{1/2}x_k$, $\|P^{1/2}x_k\|_2$, decreases by at least a factor of γ every time step k, and hence robustness with respect to time varying perturbations is guaranteed. Note that if both M is real, and $\mathbf{\Delta} \subset \mathbf{R}^{m\times m}$, then by using an argument similar to that in Theorem 9.10, the matrix P, if it exists, can be chosen to be real.

Using Theorem 4.10 and the fact that in some instances (when $2S + F \le 3$), μ and the upper bound are always equal, we can establish necessary and sufficient conditions for a pair to be quadratically stable, in terms of a scaled state space test, and/or a scaled $\mathcal{H}_\infty$ norm test. Several cases are outlined below, along with a chain of equivalences which produces the result. As usual, define the transfer function $G(z)$ as

$$G(z) := M_{22} + M_{21}(zI - M_{11})^{-1}M_{12}.$$

Note that this is the transfer function of the linear system that the perturbation Δ_k "sees".

11.1. *Real state space data,* 1 *full real perturbation*

Suppose that $M \in \mathbf{R}^{(n+m)\times(n+m)}$, and that $\mathbf{\Delta} = \mathbf{R}^{m\times m}$. Assume that $\bar{\sigma}(M_{22}) < 1$. For any $P \in \mathbf{C}^{m\times n}$ with $P = P^* > 0$, let

$$M^P := \begin{bmatrix} P^{1/2}M_{11}P^{-1/2} & P^{1/2}M_{12} \\ M_{21}P^{-1/2} & M_{22} \end{bmatrix}.$$

Also, define

$$\mathbf{\Delta}_R := \{\mathrm{diag}\,[\Delta_1, \Delta_2] : \Delta_1 \in \mathbf{R}^{n\times n}, \Delta_2 \in \mathbf{R}^{m\times m}\}.$$

Then, using Theorem 4.3 and the results from Sections 9.7 and 10, the following statements are equivalent.

(1) There exists $P \in \mathbf{C}^{n\times n}$, $P = P^* > 0$ such that $\max_{\Delta \in \mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}[P^{1/2}\mathcal{S}(M, \Delta)P^{-1/2}] < 1$.

(2) There exists $P \in \mathbf{R}^{n\times n}$, $P = P^T > 0$ such that $\max_{\Delta \in \mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}[P^{1/2}\mathcal{S}(M, \Delta)P^{-1/2}] < 1$.

(3) $\inf_{\substack{P \in \mathbf{R}^{n\times n} \\ P = P^T > 0}} \max_{\Delta \in \mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}[\mathcal{S}(M^P, \Delta)] < 1.$

(4) $\inf_{\substack{P \in \mathbf{R}^{n\times n} \\ P = P^T > 0}} \mu_{\mathbf{\Delta}_R}(M^P) < 1.$

(5) $\inf_{\substack{P \in \mathbf{R}^{n\times n} \\ P = P^T > 0}} \inf_{d_1 > 0}$

$$\bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1} & 0 \\ 0 & I_m \end{bmatrix} M^P \begin{bmatrix} \frac{1}{\sqrt{d_1}} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(6) $\inf_{\substack{P \in \mathbf{R}^{n\times n} \\ P = P^T > 0}} \inf_{d_1 > 0}$

$$\bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1}\,P^{1/2} & 0 \\ 0 & I_m \end{bmatrix} M \begin{bmatrix} \frac{1}{\sqrt{d_1}}P^{-1/2} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(7) $\inf_{\substack{P \in \mathbf{R}^{n\times n} \\ P = P^T > 0}}$

$$\bar{\sigma}\left(\begin{bmatrix} P^{1/2} & 0 \\ 0 & I_m \end{bmatrix} M \begin{bmatrix} P^{-1/2} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(8) $\rho(M_{11}) < 1$ and $\|G\|_\infty < 1$.

The main point here is that the uncertain system is quadratically stable with respect to full block, norm bounded, real perturbations (Condition 1) if and only the $\mathcal{H}_\infty$ norm of the transfer function that the perturbation sees is less than one

(Condition 8). Conditions 2–7 are intermediate steps which link the two conditions together. This same style is used in Sections 11.2–11.4.

11.2. *Complex state space data, 1 full complex perturbation*

Suppose that $M \in \mathbf{C}^{(n+m)\times(n+m)}$, and that $\Delta = \mathbf{C}^{m\times m}$. Assume that $\bar{\sigma}(M_{22}) < 1$. For any $P \in \mathbf{C}^{n\times n}$ with $P = P^* > 0$, let

$$M^P := \begin{bmatrix} P^{1/2}M_{11}P^{-1/2} & P^{1/2}M_{12} \\ M_{21}P^{-1/2} & M_{22} \end{bmatrix}$$

Also, define

$$\mathbf{\Delta}_C := \{\mathrm{diag}\,[\Delta_1, \Delta_2] : \Delta_1 \in \mathbf{C}^{n\times n}, \Delta_2 \in \mathbf{C}^{m\times m}\}.$$

Then, using Theorem 4.3 and the results from Sections 9.1 and 10, the following statements are equivalent:

(1) There exists $P \in \mathbf{C}^{n\times n}$, $P = P^* > 0$ such that $\max_{\Delta\in\mathbf{B}_\Delta} \bar{\sigma}[P^{1/2}\mathcal{S}(M, \Delta)P^{-1/2}] < 1$.

(2) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \max_{\Delta\in\mathbf{B}_\Delta} \bar{\sigma}[\mathcal{S}(M^P, \Delta)] < 1$.

(3) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \mu_{\mathbf{\Delta}_C}(M^P) < 1$.

(4) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{d_1>0}$

$$\bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1} & 0 \\ 0 & I_m \end{bmatrix} M^P \begin{bmatrix} \frac{1}{\sqrt{d_1}} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(5) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{d_1>0}$

$$\bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1}\,P^{1/2} & 0 \\ 0 & I_m \end{bmatrix} M \begin{bmatrix} \frac{1}{\sqrt{d_1}}P^{-1/2} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(6) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}}$

$$\bar{\sigma}\left(\begin{bmatrix} P^{1/2} & 0 \\ 0 & I_m \end{bmatrix} M \begin{bmatrix} P^{-1/2} & 0 \\ 0 & I_m \end{bmatrix}\right) < 1.$$

(7) $\rho(M_{11}) < 1$ and $\|G\|_\infty < 1$.

Some interesting connections between different notions of stability can be made at this point. To do so, consider the definition of robust stability given below:

Definition 11.2. The pair $(M, \mathbf{\Delta})$ is robustly stable if

$$\max_{\Delta\in\mathbf{B}_\Delta} \rho(\mathcal{S}(M, \Delta)) < 1.$$

Recall the example in Section 4, which demonstrated an application of the Main Loop theorem. In that example, LFT arguments were given to prove that the pair $(M, \mathbf{C}^{m\times m})$ is robustly stable if and only if $\|G\|_\infty < 1$, where $G(z) = M_{22} + M_{21}(zI_n - M_{11})^{-1}M_{12}$. That result, along with Section 11.1 and this section combine to form the following theorem (Willems, 1973; Popov, 1962; Khargonekar *et al.*, 1990).

Theorem 11.3. Suppose that $M \in \mathbf{R}^{(n+m)\times(n+m)}$, with $\bar{\sigma}(M_{22}) < 1$. Define $G(z) := M_{22} + M_{21}(zI_n - M_{11})^{-1}M_{12}$. Then, the conditions

(1) the pair $(M, \mathbf{R}^{m\times m})$ is quadratically stable;
(2) the pair $(M, \mathbf{C}^{m\times m})$ is quadratically stable;
(3) the pair $(M, \mathbf{C}^{m\times m})$ is robustly stable;
(4) $\rho(M_{11}) < 1$, and $\|G\|_\infty < 1$;

are equivalent.

It is important to note that Conditions (1) and (3) become incomparable (neither implies the other) when the perturbation set becomes structured (Packard and Doyle, 1990; Rotea *et al.*, 1991).

11.3. *Complex state space data, 1 repeated complex perturbation*

Suppose $M \in \mathbf{C}^{(n+m)\times(n+m)}$ and $\Delta = \{\delta I_m : \delta \in \mathbf{C}\}$. Let $\mathbf{\Delta}_S$ be defined as

$$\mathbf{\Delta}_S := \{\mathrm{diag}\,[\Delta_1, \delta_2 I_m] : \Delta_1 \in \mathbf{C}^{n\times n}, \delta_2 \in \mathbf{C}\}.$$

Then, using Theorem 4.3 and the results from Sections 9.4 and 10, the following statements are equivalent:

(1) There exists $P \in \mathbf{C}^{n\times n}$, $P = P^* > 0$ such that $\max_{\substack{\delta_2\in\mathbf{C}\\ |\delta_2|\le 1}} \bar{\sigma}[P^{1/2}\mathcal{S}(M, \delta_2 I_m)P^{-1/2}] < 1$.

(2) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \max_{\substack{\delta_2\in\mathbf{C}\\ |\delta_2|\le 1}} \bar{\sigma}[\mathcal{S}(M^P, \delta_2 I_m)] < 1$.

(3) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \mu_{\mathbf{\Delta}_S}(M^P) < 1$.

(4) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{\substack{D_2\in\mathbf{C}^{m\times m}\\ D_2=D_2^*>0}}$

$$\bar{\sigma}\left(\begin{bmatrix} I_n & 0 \\ 0 & D_2^{1/2} \end{bmatrix} M^P \begin{bmatrix} I_n & 0 \\ 0 & D_2^{-1/2} \end{bmatrix}\right) < 1.$$

(5) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{\substack{D_2\in\mathbf{C}^{m\times m}\\ D_2=D_2^*>0}}$

$$\left(\begin{bmatrix} P^{1/2} & 0 \\ 0 & D_2^{1/2} \end{bmatrix} M \begin{bmatrix} P^{-1/2} & 0 \\ 0 & D_2^{-1/2} \end{bmatrix}\right) < 1.$$

(6) $\rho(M_{11}) < 1$ and $\inf_{\substack{D_2\in\mathbf{C}^{m\times m}\\ D_2=D_2^*>0}} \|D_2^{1/2}GD_2^{-1/2}\|_\infty < 1$.

In this section, the matrix $\mathcal{S}(M, \delta I_m)$ is a

rational function of the scalar, complex parameter δ. We have shown that quadratic stability with respect to such a parameter can be ascertained by determining if the convex set

$$\left\{X = \begin{bmatrix} P & 0 \\ 0 & D_2 \end{bmatrix} : P \in \mathbf{C}^{n\times n}, D_2 \in \mathbf{C}^{m\times m}, X = X^* > 0, M^*XM - X < 0\right\},$$

is nonempty.

11.4. *Complex state space data, 2 complex full blocks*

Suppose that $M \in \mathbf{C}^{(n+m)\times(n+m)}$, and $\mathbf{\Delta}$ is

$$\mathbf{\Delta} := \{\mathrm{diag}\,[\Delta_1, \Delta_2] : \Delta_i \in \mathbf{C}^{m_i\times m_i}\} \subset \mathbf{C}^{m\times m},$$

and let

$$\mathbf{\Delta}_C := \{\mathrm{diag}\,[\Delta_0, \Delta] : \Delta_0 \in \mathbf{C}^{n\times n}, \Delta \in \mathbf{\Delta}\}.$$

Then, using Theorem 4.3 and the results from Sections 9.3 and 10, the following statements are equivalent.

(1) There exists $P \in \mathbf{C}^{n\times n}$, $P = P^* > 0$ such that $\max_{\Delta\in\mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}[P^{1/2}\mathcal{S}(M, \Delta)P^{-1/2}] < 1$.

(2) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \max_{\Delta\in\mathbf{B}_{\mathbf{\Delta}}} \bar{\sigma}[\mathcal{S}(M^P, \Delta)] < 1$.

(3) $\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \mu_{\mathbf{\Delta}_C}(M^P) < 1$.

(4) $$\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{d_1,d_2>0} \bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1}\,I_n & 0 & 0 \\ 0 & \sqrt{d_2}\,I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \times M^P \begin{bmatrix} \frac{1}{\sqrt{d_1}} I_n & 0 & 0 \\ 0 & \frac{1}{\sqrt{d_2}} I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}\right) < 1.$$

(5) $$\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{d_1,d_2>0} \bar{\sigma}\left(\begin{bmatrix} \sqrt{d_1}\,P^{1/2} & 0 & 0 \\ 0 & \sqrt{d_2}\,I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \times M \begin{bmatrix} \frac{1}{\sqrt{d_1}} P^{-1/2} & 0 & 0 \\ 0 & \frac{1}{\sqrt{d_2}} I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}\right) < 1.$$

(6) $$\inf_{\substack{P\in\mathbf{C}^{n\times n}\\ P=P^*>0}} \inf_{d_2>0} \bar{\sigma}\left(\begin{bmatrix} P^{1/2} & 0 & 0 \\ 0 & \sqrt{d_2}\,I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix} \times M \begin{bmatrix} P^{-1/2} & 0 & 0 \\ 0 & \frac{1}{\sqrt{d_2}} I_{m_1} & 0 \\ 0 & 0 & I_{m_2} \end{bmatrix}\right) < 1$$

(7) $\rho(M_{11}) < 1$ and

$$\inf_{d_2>0} \left\| \begin{bmatrix} \sqrt{d_2}\,I_{m_1} & 0 \\ 0 & I_{m_2} \end{bmatrix} G \begin{bmatrix} \frac{1}{\sqrt{d_2}} I_{m_1} & 0 \\ 0 & I_{m_2} \end{bmatrix} \right\| < 1.$$

Hence quadratic stability with respect to two full complex blocks of uncertainty is equivalent to an optimally scaled small gain condition. Note that for any $\alpha > 0$,

$$\left\{ d_2 > 0 : \left\| \begin{bmatrix} \sqrt{d_2}\,I_{m_1} & 0 \\ 0 & I_{m_2} \end{bmatrix} G \begin{bmatrix} \frac{1}{\sqrt{d_2}} I_{m_1} & 0 \\ 0 & I_{m_2} \end{bmatrix} \right\|_\infty < \alpha \right\},$$

is either empty or is a convex set.

11.5. *Conclusions*

Some of these results are well-known, and available in the literature, although the treatment here is more unified. The results relating quadratic stability and $\|\cdot\|_\infty$ and for full block perturbations (Sections 11.1 and 11.2) are proven for SISO systems in Popov (1962) and Willems (1973) and for MIMO systems in Khargonekar *et al.* (1990). The results for two complex blocks is from Packard and Doyle (1990) while the result for a single complex repeated scalar perturbation, is, to our knowledge, new. Similar results are easily derived for continuous-time systems, using a bilinear transform. By defining

$$\mathcal{B} := \begin{bmatrix} I_n & \sqrt{2}\,I_n \\ \sqrt{2}\,I_n & I_n \end{bmatrix},$$

and noting that

$$\begin{aligned} A^*P + PA < 0 &\leftrightarrow P^{-1/2}A^*P^{1/2} + P^{1/2}AP^{-1/2} < 0 \\ &\leftrightarrow \bar{\sigma}(\mathcal{S}(\mathcal{B}, P^{1/2}AP^{-1/2})) < 1 \\ &\leftrightarrow \bar{\sigma}(P^{1/2}\mathcal{S}(\mathcal{B}, A)P^{-1/2}) < 1, \end{aligned}$$

the results relating $\|\cdot\|_\infty$ (possibly a scaled norm, as in Sections 11.4 and 11.3) and quadratic stability can be derived in the same manner.

12. μ-SYNTHESIS VIA OPTIMALLY SCALED LFTS

This paper so far has only considered μ-analysis. The problem of μ-synthesis is much more difficult, and will be discussed in this section. In general, μ-synthesis methods have focused on minimizing μ of some rational matrix over stabilizing controllers using the frequency-domain upper bound (FDUB) and have been successfully used in many applications. Nevertheless, the theoretical basis for μ-synthesis is much weaker than for μ-analysis. This section will consider a μ-synthesis problem involving only constant matrices, to explore the potential difficulties in a simple setting. For an introduction to μ-synthesis in the rational case, see Balas *et al.* (1991).

Suppose that a matrix M depends on a free parameter Q. How can Q be found so as to minimize $\mu_{\mathbf{\Delta}}(M)$? In this section we consider this problem when M depends on a free matrix Q in a linear fractional manner, and we attempt to minimize the upper bound for $\mu_{\mathbf{\Delta}}(M)$, rather than $\mu_{\mathbf{\Delta}}(M)$ itself. This problem is first reduced to an affine, rather than linear fractional, transformation, and then partially solved using a elementary extension to matrix dilation theory (Davis *et al.*, 1982; Power, 1982). In the lemmas to follow, $\mathbf{F}$ denotes either the real or complex field.

Lemma 12.1. Let $R \in \mathbf{F}^{n\times n}$, $U \in \mathbf{F}^{n\times r}$, $T \in \mathbf{F}^{t\times r}$, and $V \in \mathbf{F}^{t\times n}$, where $r, t \le n$. Let $\mathbf{Z} \subset \mathbf{F}^{n\times n}$ be a prescribed set of positive definite matrices. Then

$$\inf_{\substack{Q\in\mathbf{F}^{r\times t}, Z\in\mathcal{Z} \\ \det(I-TQ)\neq 0}} \bar{\sigma}[Z^{1/2}(R + UQ(I - TQ)^{-1}V)Z^{-1/2}]$$

$$= \inf_{\tilde{Q}\in\mathbf{F}^{r\times t}, Z\in\mathcal{Z}} \bar{\sigma}[Z^{1/2}(R + U\tilde{Q}V)Z^{-1/2}].$$

Proof. For any $T \in \mathbf{F}^{t\times r}$, the closure of the set

$$\{Q(I - TQ)^{-1} : Q \in \mathbf{F}^{r\times t}, \det(I - TQ) \neq 0\}.$$

is all of $\mathbf{F}^{r\times t}$, which shows that the infimums are the same.

Hence, in order to solve general linear fractional transformation optimization problems, only affine transformations need be considered. We also assume (without loss in generality) that U is full column rank, and that V is full row rank. The first lemma addresses the unscaled problem, and comes from Davis *et al.* (1982) and Power (1982).

Lemma 12.2. Let R, U, V, be given as above. Suppose $U_\perp \in \mathbf{F}^{n\times(n-r)}$ and $V_\perp \in \mathbf{F}^{(n-t)\times n}$ are chosen such that $[U\ U_\perp]$, $\begin{bmatrix} V \\ V_\perp \end{bmatrix}$ are both invertible, and that

$$U^*U_\perp = 0_{r\times(n-r)}, VV_\perp^* = 0_{t\times(n-t)}.$$

Let $\alpha > 0$. Then

$$\inf_{Q\in\mathbf{F}^{r\times t}} \bar{\sigma}[(R + UQV)] < \alpha,$$

if and only if

$$\begin{aligned} \lambda_{\max}[V_\perp(R^*R - \alpha^2 I)V_\perp^*] &< 0, \\ \lambda_{\max}[U_\perp^*(RR^* - \alpha^2 I)U_\perp] &< 0. \end{aligned} \quad (12.1)$$

The next lemma partially answers the synthesis question when similarity scalings are included. The proof is in Doyle (1985a) and Packard *et al.* (1992).

Lemma 12.3. Let R, U, V, $U_\perp$ and $V_\perp$ be given as above. Let $\alpha > 0$ and $\mathbf{Z} \subset \mathbf{F}^{n\times n}$ be a given set of positive definite, Hermitian matrices. Then

$$\inf_{\substack{Q\in\mathbf{F}^{r\times t} \\ Z\in\mathbf{Z}}} \bar{\sigma}[Z^{1/2}(R + UQV)Z^{-1/2}] < \alpha,$$

if and only if there is a $Z \in \mathbf{Z}$ such that

$$\lambda_{\max}[V_\perp(R^*ZR - \alpha^2 Z)V_\perp^*] < 0, \quad (12.2)$$

and

$$\lambda_{\max}[U_\perp^*(RZ^{-1}R^* - \alpha^2 Z^{-1})U_\perp] < 0. \quad (12.3)$$

Note that the condition imposed on Z in equation (12.2), is convex, therefore, if the set $\mathbf{Z}$ is itself convex, determining solutions of equation (12.2) is a convex feasibility problem. Similarly, the condition imposed on Z^{-1} in equation (12.3) is convex in Z^{-1}, so if the set $\mathbf{Z}^{-1}$ is convex, this is also a convex feasibility problems is exploited in Packard *et al.* (1991, 1992) where some robust control prob-
(1991, 1992) where some robust control prob-
lems are formulated, and recast as convex optimizations, using this scaled linear fractional transformation approach. Unfortunately, the complete problem, which involves both Z and Z^{-1} conditions, is more difficult, and at the moment, unsolved. In some special cases, it may be possible to obtain computable necessary and sufficient conditions. For instance, if

$$\mathbf{Z} = \left\{\begin{bmatrix} z_1 I_{n_1} & 0 \\ 0 & I_{n_2} \end{bmatrix} : z_1 > 0\right\},$$

then both conditions define open intervals in the real line, and it is easy to check if these intervals intersect (moreover, the intersection is either empty or convex). More generally though, the set of "good Zs" may be a disconnected set. Specifically, given matrices R, U and V, and

$\alpha > 0$, let $\mathbf{Z}_{\text{good}}$ be

$$\mathbf{Z}_{\text{good}}(\alpha) := \left\{ Z \in \mathbf{Z}: \inf_{Q \in \mathcal{M}(\mathbf{C})} \bar{\sigma}(Z^{1/2}(R + UQV)Z^{-1/2}) < \alpha \right\}.$$

It is this set, $\mathbf{Z}_{\text{good}}(\alpha)$, which may be disconnected. In particular, let $\alpha := 1$, and

$$\mathbf{Z} = \{\text{diag}\,[z_1, z_2, 1]: z_1 > 0, z_2 < 0\},$$

and define matrices

$$R = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 10\sqrt{2} \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

$$U = V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 \\ 0 & 1 \end{bmatrix}$$

Applying the formulae, we have that $Z \in \mathbf{Z}_{\text{good}}$ if and only if $z_1 > 0$, $z_1 + z_2 > 2$, and $\frac{1}{z_1} + \frac{1}{z_2} > 200$. Clearly, the region of good Zs in the z_1–z_2 plane consists of two slivers near the axis, which is not a connected set. Unlike the analysis problem the level sets of scalings in the synthesis problem are not convex. We are currently investigating the implications of this property.

13. SUMMARY OF SOME RELATED WORK

This section outlines some work related to this paper, beginning with a brief history of the early development of the μ theory. This outline is not intended to be exhaustive or complete, but simply to touch on a few of the topics nearest to this paper that were not considered in detail. LMIs are discussed as potentially unifying theoretical and computational tools. The relationship between μ and quadratic vs L_1 notions of robust performance and robust stability is considered next, followed by μ with mixed real and complex perturbations. The section ends with model validation and generalizations of μ.

13.1. *History of early work*

In this section, we will briefly review the ideas that most influenced the original development of the μ theory. These remarks are drawn mainly from earlier papers (Doyle, 1982; Doyle *et al.*, 1982; Fan *et al.*, 1991), but are repeated here for the convenience of the reader.

An obvious influence on the developement of the μ theory was the work in so-called Robust Multivariable Control Systems from the late 1970s, (IEEE, 1981) which in turn drew heavily on earlier work in stability analysis (e.g. Zames, 1965; Desoer and Vidyasagar, 1975; Willems, 1971b; Safonov, 1980), particularly the small gain and circle theorems. These theorems established sufficient conditions for stability of nonlinear components connected in feedback. The emphasis in the early robustness work was on small gain type conditions involving singular values that were both necessary and sufficient for stability of sets of linear systems involving a single norm bounded but otherwise unconstrained perturbation. Another emphasis for much of the robustness theory was on using singular value plots as a means of generalizing Bode magnitude plots to multivariable systems.

While methods based on singular values were gaining in popularity, it became evident that their assumption of unstructured uncertainty was too crude for many applications. Furthermore, the problem of robust performance was not adequately treated. Freudenberg *et al.* (1982) studied these issues using differential sensitivity and suggested that something more than singular values was needed. It was a natural step to introduce structured uncertainty of the type considered in this paper (see Safonov (1978) for an early treatment). The so-called conservativeness of singular values was based on the fact that the unscaled bounds $\rho(M) \le \mu_\Delta(M) < \bar{\sigma}(M)$ could be arbitrarily far off, and research was begun to provide improved estimates of μ, with an initial focus on the nonrepeated, complex case ($S = 0$).

It was obvious that the sharper bounds

$$\max_{Q \in \mathbf{Q}} \rho(QM) : \max_{\Delta \in \mathbf{B\Delta}} \rho(\Delta M) = \mu_\Delta(M) \le \inf_{D \in \mathbf{D}} \bar{\sigma}(D^{1/2} M D^{-1/2}), \quad (13.1)$$

could help alleviate the conservativeness somewhat. The upper bound is similar to the multiplier methods that were used in nonlinear stability analysis to reduce the conservativeness of small gain type methods (Willems, 1971b) but the use of both upper and lower bounds, and the questions of how close the bounds were and how to efficiently compute them were new and open. As we saw in Section 6, the equality of the lower bound and μ is relatively straightforward and not surprising. What is remarkable, even in retrospect, is that the upper bound is often close to μ and is in fact equal to μ for certain simple block structures.

There was substantial numerical evidence for the upper bound results before they were proven. Engineers at Honeywell's Systems and

Research Center, particularly Joe Wall, began routinely using a simple generalization of Osborne's routine (Osborne, 1960) to approximate the upper bound in (13.1) and gradient search methods to find a local maximum for the lower bound. Osborne's algorithm minimizes the Frobenius norm rather than the maximum singular value, and the scalings produced can be used to approximate the upper bound. The consistent closeness of the bounds, usually within a few percent, suggested that there was a deeper connection between the bounds. Ironically, minimizing the Frobenius norm remains the cheapest method of approximating the upper bound. Safonov (1982) suggested a somewhat less general approximation to the upper bound based on Perron eigenvectors which is comparable to Osborne in speed and accuracy.

While the μ framework arises naturally in studying robust stability with structured uncertainty, the use of μ to treat directly the problem of robust performance with structured uncertainty was first explicitly noted in Doyle *et al.* (1982). As noted above, this is a consequence of the intimate connection between μ and LFTs (Doyle, 1985b; Packard, 1987). In retrospect, it is clear that Redheffer (1959, 1960) had developed the foundation of this connection in his work on LFTs in the late 1950s. In fact, as noted earlier, Redheffer had even proven that the upper bound in (13.1) was an equality for the case where $S = 0$ and $F = 2$. While Redheffer's results were not well-known in the control community until the μ theory was already well-developed, the rediscovery of his work has since had an important influence, not only on the further development of μ but in other areas as well (e.g. see Doyle *et al.*, 1989).

13.2. *Linear matrix inequalities*

We have seen in this paper how LMIs arise naturally in both μ analysis and synthesis in the computation of upper bounds. The general LMI problem involves sets of the form

$$\mathscr{X} = \left\{ \begin{matrix} \operatorname{diag}[X_1, \ldots, X_S, x_1 I, \ldots, x_F I] : \\ X_i \in \mathbf{C}^{r_i \times r_i}, X_i = X_i^*, x_j \in \mathbf{R} \end{matrix} \right\}, \quad (13.2)$$

and a list of matrices A_i, B_i, C_i, D_i. The simplest general LMI problem is to determine whether there exists $X \in \mathscr{X}$ such that

$$A_i^* X A_i - B_i^* X B_i + X C_i + C_i^* X + D_i < 0 \quad \forall i.$$

Depending on the particular problem, the $<$ may be a $\leq$. It is easy to see that these inequality conditions produce a set of solutions which are convex, which makes LMIs attractive computationally. (In the synthesis problem in Section 12, there are additional constraints that destroy convexity.) This is a decision problem; the answer is yes or no. Sometimes, however, the A_i, B_i, C_i, and D_i are functions of a real, positive parameter α, and we want to know, for example, what is the largest α for which there is no solution. Typically this involves an iteration on α, and consequently, answering the decision question many times.

Recall that the upper bound for μ can be rewritten as an LMI of the form

$$\exists X > 0 : M^* X M - \beta^2 X < 0. \quad (13.3)$$

It has recently been shown that a number of other problems can be reduced to solving LMIs. In Wang *et al.* (1991) balanced truncation model reduction is extended to uncertain LFT systems, with similar extensions of the parametrization of all stabilizing controllers in Lu *et al.* (1991). The LFT/LMI machinery not only extends the standard results in important ways, it simplifies the proofs, often substantially. Exciting new developments in handling real parametric uncertainty (Young *et al.*, 1991) and model validation (Newlin and Smith, 1991) will be outlined in subsequent subsections. In all cases, LMIs play a central role in computation of solutions. We believe that LMIs will replace Lyapunov and Riccati equations, which are both special cases of LMIs, as the central computational problems in robust control.

The problem of solving LMIs can be viewed in a number of ways, from solving a set of linear equalities to minimizing the eigenvalues of a Hermitian matrix function (Beck, 1991). One of the goals of our current research is to develop fast, reliable algorithms for solving LMIs which are comparable to what is available for solving Riccati and Lyapunov equations. Several researchers have already begun looking at this question.

One approach to solving LMIs is to convert them to eigenvalue optimization problems which results in convex, non-differentiable functions for which numerous optimization methods have been developed. Boyd and Yang (1989) compare the efficiency of two convex programming algorithms, Kelley's cutting-plane algorithm and Shor's subgradient algorithm. Boyd and Yang find Kelley's cutting-plane algorithm to be most effective of the two, since it converges in the fewest number of iterations. An alternative convex programming method which has been used for LMI problems is the ellipsoid method, which is also a cutting-plane method. Although these methods are easy to implement, they are generally too slow to warrant considerable attention. Overton (1990) studies the optimality

conditions, and develops quadratically convergent algorithms for LMIs.

Recently, interior point methods have been applied to LMI problems with favorable results. Interior and exterior point methods are used to convert constrained minimization problems to differentiable, unconstrained minimization problems, to which optimization algorithms such as Newton's method are applied (Nesterov and Nemirovsky, 1989, 1990). Jarre (1991) has used an interior point algorithm for a problem similar to LMIs which required substantially fewer iterations than does the cutting-plane algorithm used by Boyd and Yang. A similar approach is taken by Boyd and El Ghaoui (1993). We are currently investigating alternative functions for solving LMIs using both interior and exterior point methods (Beck, 1991).

13.3. μ, $\mathcal{Q}$, *and* L_1

We have considered several different measures of robust stability and performance in Section 10 from SSμ to the SSUB. We will concentrate on these two measures, and compare them briefly with another very important measure that has emerged in the L_1 theory of robust performance with structured uncertainty. Space constraints preclude a review of the L_1 theory, which has undergone a dramatic and impressive development in the last five years in the work of Khammash and Pearson (1991) and Dahleh and Khammash (1991) and references therein. For simplicity, we will refer to the SSμ test as μ and the SSUB upper bound as $\mathcal{Q}$ (since it is directly related to quadratic stability), and focus our attention on the robust performance problem, which clearly includes robust stability as a special case.

The μ, $\mathcal{Q}$, and L_1 tests all guarantee robust performance, but with different assumptions about perturbations and the norm used for measuring the performance objective. The μ and $\mathcal{Q}$ theories are used for L_2 induced norms, while the L_1 theory is used for L_∞ induced norms. A second distinction is that the μ theory treats LTI perturbations, and the $\mathcal{Q}$ and L_1 handle Nonlinear and Time-Varying perturbations (NTV). This is summarized in the table below.

	LTI	NTV
L_2	μ	$\mathcal{Q}$
L_∞	μ	L_1

The cases on the diagonal, L_2/LTI and L_∞/NTV, are both necessary and sufficient for robust performance. The L_∞/LTI case is necessary and sufficient for robust stability, but the robust performance question is open. The $\mathcal{Q}$ case (L_2/NTV) is sufficient for robust performance, and recent results, obtained independently using very different methods by Shamma and Megretskii, suggest that it is necessary as well. Recall that in general, μ is computed using bounds, but that $\mathcal{Q}$ involves solving LMIs, so is attractive computationally. L_1 is also easy to compute, involving only the evaluation of L_1 norms and finding the spectral radius of a positive matrix (Khammash and Pearson, 1991).

As a final comparison, it can be easily shown that the tests are ordered, with

$$\mu \leq \mathcal{Q} \leq L_1. \tag{13.4}$$

The interpretation of (13.4) for a given system is that if the $\mathcal{Q}$ test passes, the μ test must pass, and similarly for L_1 and $\mathcal{Q}$. It was shown above that $\mu \leq \mathcal{Q}$. The inequality $\mathcal{Q} \leq L_1$ follows from the equivalence of the SSUB and the FDCD problems, the fact that the L_1 norm of a convolution kernal is greater than the H_∞ norm of its transform, and the results in Khammash and Pearson (1991). The inequalities are typically strict and it is possible for the gaps to be arbitrarily large.

It is not clear exactly what the implications of these results are for control design or for further research. Clearly there is a need for more refined results, and the ability to both combine LTI and NTV uncertainty and exploit additional structure such as the slowly-varying nature of some perturbations. The results in Safonov (1984), Packard and Teng (1990) and Packard and Zhou (1989) suggest how this might be done in the LFT/μ/$\mathcal{Q}$ framework, but much more work is needed. We also need more precise modeling and ID methods to exploit the detailed structure of the uncertainty in our models.

If one accepts $\mathcal{Q}$ as the measure of robust performance, a rich theory can be developed, with generalizations to uncertain systems of the conventional theories of robust stability and performance, balanced realizations and model reduction (Wang *et al.*, 1991) stabilization (Lu *et al.*, 1991) and model validation (Newlin and Smith, 1991). It is not surprising that the easiest generalizations of standard results to uncertain LFT systems is done using the $\mathcal{Q}$ framework. Indeed, most of the standard results rely on $\mathcal{Q}$ machinery, but since μ and $\mathcal{Q}$ are the same for these simple block structures, we are less aware of the distinction. Once we begin extending our results to systems with uncertainty, the distinction becomes significant. Of course, a key feature of the $\mathcal{Q}$ theory is that computation involves solving LMIs.

13.4. *μ with real perturbations*

In recent years a great deal of interest has arisen with regard to robustness problems involving parametric uncertainty. These problems involve uncertain parameters that are not only norm bounded, but also constrained to be real. Robustness problems involving parametric uncertainty can be reformulated as μ problems where the block structured uncertainty description is now allowed to contain both real and complex blocks. This mixed μ problem can have fundamentally different properties from the complex μ problem studied in this paper (where the block structured uncertainty description contains only complex blocks) and these properties have important implications for computation. In this section we give a brief review of some recent results in this area.

It is now well known that real μ problems can be discontinuous in the problem data (see Barmish *et al.*, 1989). As well as adding computational difficulties to the problem this sheds serious doubt on the usefulness of real μ as a robustness measure in such cases, since the system model is always a mathematical abstraction from the real world, and is computed to finite precision. However it is shown in Packard and Pandey (1991) that mixed μ problems containing some complex uncertainty are, under some mild assumptions, continuous in the problem data (whereas purely real μ problems are not). This is reassuring from an engineering viewpoint since one is usually interested in robust performance problems (which therefore contain at least one complex block) or robust stability problems with some unmodeled dynamics, which are naturally covered with complex uncertainty. Thus in problems of engineering interest, the potential discontinuity of mixed μ should not arise.

Recent results in Rohn and Poljak (1992) show that a special case of computing μ with real perturbations only is NP complete. While these results do not apply to the complex only case, it is certainly true that the general mixed problem is NP hard as well. These results strongly suggest that it is futile to pursue exact methods for computing μ in the purely real or mixed case for even moderate (less than 100) numbers of real perturbations, unless one is prepared not only to solve the real μ problem but also to make fundamental contributions to the theory of computational complexity. Furthermore, it may be that even approximate methods must have worst-case combinatoric complexity (Demmel, 1992).

These results do not mean, however, that "practical" algorithms are not possible, where "practical" means avoiding combinatoric (non-polynomial) growth in computation with the number of parameters for all of the problems which arise in engineering applications. Practical algorithms for other NP hard problems exist and typically involve approximation, heuristics, branch-and-bound, or local search. Results presented in Young *et al.* (1992) strongly suggest that an intelligent combination of all these techniques can yield a practical algorithm for the mixed problem.

Upper and lower bounds for mixed μ have recently been developed, and they take the form of generalizations of the bounds for the complex μ problem presented here (i.e. by applying the mixed μ bounds to complex μ problems one recovers the standard complex μ bounds). The upper bound was presented in Fan *et al.* (1991) and involves minimizing the eigenvalues of a Hermitian matrix. This can also be recast as a singular value minimization which involves additional scaling parameters to the complex μ upper bound. It is shown in Young and Doyle (1990) that the mixed μ problem can be recast as a real eigenvalue maximization and that this in turn can be tackled via a power algorithm, giving a lower bound for mixed μ. A practical computation scheme for these bounds has recently been developed (Young *et al.* (1992)) and will be available shortly in a test version in conjunction with the μ-Tools toolbox (Balas *et al.*, 1991).

The quality of these bounds, and their computational requirements as a function of problem size, are explored in Young *et al.* (1991). While the bounds are usually accurate enough for engineering purposes, in a significant number of cases of interest, they are not. This is in contrast with the purely complex nonrepeated case, where no examples of problems with large gaps have been found. The use of branch-and-bound schemes to improve upon existing bounds has been suggested by several authors (see Balakrishnan *et al.*, 1991; Sideris and Peña, 1989, 1990; de Gaston and Safonov, 1988 and references therein). There are some important issues and tradeoffs to be considered in implementing such a scheme, which can greatly impact the performance. A selection of results from a fairly extensive numerical study of these issues is presented in Young *et al.* (1991) and a branch-and-bound scheme is proposed which should form the basis of a practical computation scheme for mixed μ. This will be further explored in Newlin *et al.* 1993).

The upper and lower bounds from complex μ theory not only serve as computational schemes, but are theoretically rich as well. Connections

between the bounds and various aspects of linear system theory have been established, and further work in this area appears to have great promise. A theoretical study of the mixed μ bounds may yield new insight as well, and this is a subject of current research. Initial results in this area are presented in Young and Doyle (1993), where it is seen that mixed μ inherits many of the (appropriately generalized) properties of complex μ, although as has already been seen, in some aspects the mixed μ problem can be fundamentally different from the complex μ problem.

Problems involving robustness properties of polynomials with coefficients perturbed by real parameters have received a great deal of attention in the literature. This type of robustness problem leads to a (real or) mixed μ problem. Several celebrated "Kharitonov-type" results have been proven for special cases of this problem, such as the "affine parameter variation" problem (see Barlett *et al.*, 1988 for example), and the solutions typically involve checking the edges or vertices of some polytope in the parameter space. It can be shown that restricting the allowed perturbation dependence to be affine leads to a real μ problem on a transfer matrix which is rank one.

The rank one mixed μ problem is studied in detail by Chen *et al.* (1991) see also the references therein. The authors develop an analytic expression for the solution to this problem, which is not only easy to compute, but has sublinear growth in the problem size. They are then able to solve several problems from the literature, noting that these problems can be treated as special cases of "rank one μ problems" and are thus "relatively easy to solve". Even the need to check (a combinatoric number of) edges is shown to be unnecessary. While many of these results were apparently well-known (Chen *et al.*, 1991) provides a direct comparison between the polynomial and μ-based approaches.

This rank one case is also studied by Young and Doyle (1990) where it is shown that for such problems μ equals its upper bound and is hence equivalent to a convex problem. This reinforces the results of Chen *et al.* (1991) and offers some insight into why the problem becomes so much more difficult when we move away from the "affine parameter variation" case to the "multilinear" or "polynomial" cases (Sideris and Peña, 1989, 1990). These correspond to μ problems which are not necessarily rank one, and hence may no longer be equal to the upper bound and so may no longer be equivalent to a convex problem. These results also underline why there are no practical algorithms based on "edge-type" theorems, as the results appear to be relevant only to a very special problem. Furthermore, even in the very special "affine parameter case" there are a combinatoric number of edges to check.

13.5. *Generalization of μ*

In this section we review an alternative formulation of μ due to Fan and Tits (1986) and use it to consider one of several possible generalizations of μ. The most important motivation for this generalization comes from the model validation problem (see Smith and Doyle (1992) and Newlin and Smith (1991) for background).

For simplicity, the Fan–Tits formulation is considered here for the full block only case ($S = 0$). For any vector or matrix A with n rows let A_i denote the rows of A corresponding to the ith block of Δ. Thus A_i has m_i rows. Also, let $P := I_n$ be the identity matrix. An alternative expression for μ is

$$\mu = \max_x \{\alpha : \alpha \|x_i\| \leq \|M_i x\| \; \forall i \in \{1, \ldots, F\}\}. \tag{13.5}$$

To see that this is equivalent to (3.3) in Definition 3.1, note that when $\det(I - M\Delta) = 0$ there is an $x \neq 0$ that satisfies $(I - \Delta M)x = 0$. This x achieves the maximum in (13.5). Conversely, any x that achieves the maximum provides a way to constructing a Δ: set Δ_i equal to the dyad that satisfies $x_i = \Delta_i M_i x$.

An LMI formulation of the upper bound follows easily from (13.5). Again, we consider the full block only case.

$$\begin{aligned} &\alpha \leq \mu \\ &\Leftrightarrow \exists x \neq 0 : \alpha^2 \|x_i\|^2 \leq \|M_i x\|^2 \; \forall i \\ &\Leftrightarrow \exists x \neq 0 : x^*(M_i^* M_i - \alpha^2 P_i^* P_i)x \geq 0 \\ &\Leftrightarrow \exists x \neq 0 : x^*(M^* D M - \alpha^2 D)x \geq 0 \quad \forall D \in \mathcal{D} \end{aligned}$$

where $\mathcal{D} := \left(D = \sum_i d_i P_i^* P_i : d_i > 0 \; \forall i\right) = \mathbf{D}$.

It follows that

$$\alpha > \mu \Leftarrow \exists D \in \mathcal{D} : M^* D M - \alpha^2 D < 0. \tag{13.6}$$

This is the same as the LMI in equation (3.11).

The generalization of μ that we will study depends on a block structure as before along with an index that specifies certain blocks as special or distinguished (Newlin and Smith, 1991). As an example, consider equation (13.5) in the case of two full blocks:

$$\mu = \max_x \{\alpha : \alpha \|x_1\| \leq \|M_1 x\| \text{ and } \alpha \|x_2\| \leq \|M_2 x\|\}.$$

Suppose the second block has been designated as special. Then the generalization is

$$\mu = \max_x \{\alpha : \alpha \|x_1\| \le \|M_1 x\| \text{ and }$$
$$\alpha^{-1} \|x_2\| \ge \|M_2 x\|\}.$$

In this example, the designation of the second block as special means that the direction of the second inequality is reversed and the scaling changed.

The lower bound for this generalization of μ, though notationally awkward, is very similar to the standard lower bound, and a power algorithm is being investigated. There is no upper bound similar to the $\bar{\sigma}(D^{1/2}MD^{-1/2})$ upper bound, but there is a generalization of the LMI above. Again consider our two block example.

$$\alpha \le \mu$$
$$\Leftrightarrow \exists x \ne 0 \begin{cases} \alpha^2 \|x_1\|^2 \le \|M_1 x\|^2 \\ \alpha^{-2} \|x_2\|^2 \ge \|M_2 x\|^2 \end{cases}$$
$$\Leftrightarrow \exists x \ne 0 : \begin{cases} x^*(M_1^* M_1 - \alpha^2 P_1^* P_1)x \ge 0 \\ x^*(M_2^* M_2 - \alpha^{-2} P_2^* P_2)x \le 0 \end{cases}$$
$$\Leftrightarrow \exists x \ne 0 : x^*(M^*DM - P^2(\alpha)D)x \ge 0 \quad \forall D \in \mathcal{D}$$

where $\mathcal{D} := (d_1 P_1^* P_1 + d_2 P_2^* P_2 : d_1 > 0; d_2 < 0)$

and $P(\alpha) = \alpha^2 P_1^* P_1 + \alpha^{-2} P_2^* P_2$.

It follows that

$$\alpha > \mu \Leftarrow \exists D \in \mathcal{D} : M^*DM - P(\alpha)^2 D < 0.$$

We see that D is just as in the case of standard LMI upper bound for μ except that we require for some blocks that $D_i < 0$ rather than $D_i > 0$. It is expected that algorithms for computing positive definition solutions to LMIs will be easily generalized to solve this problem.

Acknowledgements—The authors gratefully acknowledge helpful discussions with Blaise Morton, Sigurd Skogestad, Peter Young, Michael Fan, and Kemin Zhou and excellent comments on this paper from Carolyn Beck and Jorge Tierno. Thanks also to Matt Newlin, Sharon Laubach, Brian Clendenin, and Sonja Glavaski for useful suggestions. Financial support for this work came from the National Science Foundation (grants ECS-8909499, CTS-9057420, and ECS-8657814) AFOSR, ONR, NASA, Honeywell, Rockwell International, and the Berkeley Engineering Fund.

REFERENCES

Anderson, B., P. Agathoklis, E. Jury and M. Mansour (1986). Stability and the matrix Lyapunov equation for discrete 2-dimensional systems. *IEEE Trans. on Circuits and Sys.*, **33**, 261–267.

Bamieh, B. and M. Dahleh (1992). On robust stability with structured time-invariant perturbations. *Center for Control Engineering and Computation*, UC Santa Barbara, CCEC-92-0331.

Balakrishnan, V., S. Boyd and S. Balemi (1991). Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems. Technical report, Information Systems Laboratory, Stanford University.

Balas, G., J. Doyle, K. Glover, A. Packard and R. Smith (1991). The μ analysis and synthesis toolbox. Math Works and MUSYN.

Bartlett, A. C., C. V. Hollot and H. Lin (1988). Root locations of an entire polytope of polynomials: it suffices to check the edges. *Mathematics of Control, Signals, and Systems*, **1**, 1.

Barmish, B., P. Khargonekar, Z. Shi and R. Tempo (1989). Robustness margin need not be a continuous function of the problem data. *Systems and Control Letters*, **15**, 91–98.

Beck, C. (1991). Computational issues in solving LMIs. *Proc. of the 30th IEEE Conf. Decision and Control*, 1259–1260.

Boyd, S. and C. Desoer (1985). Subharmonic functions and performance bounds on linear time invariant feedback systems. *IMA J. of Mathematical Control and Information*, **2**, 153–170.

Boyd, S. and L. El Ghaoui (1993). Method of centers for minimizing generalized eigenvalues. *Linear Algebra and its Applications*, to appear.

Boyd, S. and Q. Yang (1989). Structured and simultaneous Lyapunov functions for system stability problems. *Int. J. Control*, **49**, 2215–2240.

Chen, M. J. and C. A. Desoer (1982). Necessary and sufficient condition for robust stability of linear distributed feedback systems. *Int. J. Control*, **35**, 255–267.

Chen, M. J. K. H. Fan and C. N. Nett (1991). The structured singular value and stability of uncertain polynomials: A missing link. ASME annual winter meeting, Atlanta, GA.

Dahleh, M. A. and M. H. Khammash (1992). Controller design for plants with structured uncertainty. *Automatica*, **29**, 37–56.

Daniel, R. W., B. Kouvaritakis and H. Latchman (1986). Principal direction alignment: a geometric framework for the complete solution to the μ problem. *IEE Proceedings*, **133**, 45–56.

Davis, C., W. Kahan and H. Weinberger (1982). Norm preserving dilations and their applications to optimal error bounds. *SIAM J. on Numerical Analysis*, **19**, 445–469.

de Gaston, R. and M. Safonov (1988). Exact calculation of the multiloop stability margin. *IEEE Trans. Aut. Control*, **33**, 156–171.

Demmel, J. (1992). The componentwise distance to the nearest singular matrix. *SIAM J. on Matrix Analysis and Applications*, **13**, 10–19.

Desoer, C. A. and M. Vidyasagar (1975). *Feedback Systems: Input–Output Properties*. Academic Press, New York.

Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proceedings*, **129**, 242–250.

Doyle, J. C. (1985a). Matrix interpolation theory and $\mathcal{H}_\infty$ performance bounds. *Proc. American Control Conf.*, Boston, MA.

Doyle, J. C. (1985b). Structured uncertainty in control system design. *IEEE Conf. on Decision and Control*, Ft. Lauderdale.

Doyle, J., K. Glover, P. Khargonekar and B. Francis (1989). State space solutions to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Trans. Aut. Control*, **34**, 831–847.

Doyle, J. C. and A. Packard. Uncertain multivariable systems from a state space perspective. *Proc. of American Control Conf.*, pp. 2147–2152.

Doyle, J., A. Packard and K. Zhou (1991). Review of LFTs, LMIs and μ. *Proc. of the 30th IEEE Conf. on Decision and Control*, pp. 1227–1232.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. Aut. Control*, **AC-26**, 4–16.

Doyle, J. C., J. Wall and G. Stein (1982). Performance and robustness analysis for structured uncertainty. *IEEE Conf. on Decision and Control*, pp. 629–636.

Dunford, N. and J. T. Schwartz (1958). *Linear Operators*. Interscience Publishers, Wiley, New York.

Fan, M. K. H. and A. L. Tits (1986). Characterization and efficient computation of the structured singular value. *IEEE Trans. Aut. Control*, **AC-31**, 734–743.

Fan, M., A. Tits and J. Doyle (1991). Robustness in the

presence of joint parametric uncertainty and unmodeled dynamics. *IEEE Trans. Aut. Control*, **36**, 25–38.

Foo, Y. K. and I. Postlethwaite (1988). Extensions of the small-μ test for robust stability. *IEEE Trans. Aut. Control*, **33**, 172–176.

Freudenberg, J. S., D. P. Looze and J. B. Cruz (1982). Robustness analysis using singular value sensitivities. *Int. J. Control*, **35**, 95–116.

Helton, W. (1988). A numerical method for computing the structured singular value. *Systems and Control Letters*, 21–26.

Special issue on linear multivariable control systems. *IEEE Trans. Aut. Control*, **26**, 621–623.

Jarre, F. (1991). An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices. Technical Report SOL 91-8, Stanford, CA.

Kato, T. (1982). *A Short Introduction to Perturbation Theory for Linear Operators*. Springer Verlag, New York.

Kharitonov, V. L. (1979). Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differential Equations*, **14**, 1483–1485.

Khargonekar P. and I. Kaminer (1991). Robust stability analysis with structure norm bounded unstable uncertainty. *Amer. Control Conf.*, pp. 2700–2701, Boston, MA.

Khammash, M. and J. B. Pearson (1991). Performance robustness of discrete-time systems with structured uncertainty. *IEEE TAC*, **AC-36**, 398–412.

Khargonekar, P., I. Petersen and K. Zhou (1990). Robust stabilization of uncertain linear systems: quadratic stability and $\mathcal{H}_\infty$ control theory. *IEEE Trans. Aut. Control*, **35**, 356–361.

Lu, W. M., K. Zhou and J. Doyle (1991). Stability and stabilization of LSI multidimensional systems. *Proc. of the 30th IEEE Conf. on Decision and Control*, 1239–1244.

Luenberger, D. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

Morton, B. and J. Doyle (1985). Private communication.

Morton, B. and R. McAfoos (1985). A mu-test for real-parameter variations. *Proc. American Control Conf.*, pp. 135–138, Boston.

Nesterov, Yu. E. and A. S. Nemirovsky (1989). *Self-Concordant Functions and Polynomial Time Methods in Convex Programming*. Academy of Science, Central Econ. and Math. Inst., Moscow, Russia.

Nesterov, Yu. E. and A. S. Nemirovsky (1990). *Optimization over Positive Semidefinite Matrices: Mathematical Background and Users Manual*. USSR Academy of Science, Central Econ. and Math. Inst., Moscow, Russia.

Newlin, M. and R. Smith (1991). Model validation and a generalization of μ. *Proc. of the 30th IEEE Conf. on Decision and Control*, pp. pp. 1257–1258.

Newlin, M. P., P. M. Young and J. C. Doyle (1993). Improving the bounds for mixed μ problems via Branch and Bound techniques. In preparation.

Osborne, E. E. (1960). On Preconditioning of Matrices. *J. Assoc. Comp. Mach.*, **7**, 338–345.

Overton, M. (1990). Large-scale optimization of eigenvalues. NYU Computer Science Department Report.

Packard, A. (1988). What's new with μ. Ph.D. Thesis, Mechanical Engineering, University of California, Berkeley.

Packard, A. and J. Whitney Balsamo (1988). A maximum modulus theorem for linear fractional transformations. *Systems and Control Letters*, **11**, 365–367.

Packard, A. and J. Doyle (1990). Quadratic stability with real and complex perturbations. *IEEE Trans. Aut. Control*, **35**, 198–201.

Packard, A., M. Fan and J. Doyle (1988). A power method for the structured singular value. *IEEE Conf. on Decision and Control*, pp. 2132–2137, Austin, TX.

Packard, A. and P. Pandey (1981). Continuity of the real/complex structured singular value. *29th Allerton Conference on Control Communication and Computing*, pp. 875–884; *IEEE Trans. Aut. Control*, to appear, March 1993.

Packard, A. and J. Teng (1990). Robust stability with time-varying perturbations. *28th Allerton Conference*, pp. 765–767.

Packard, A. and K. Zhou (1989). Improved upper bounds for the structured singular value. *28th IEEE Conference on Decision and Control*, Tampa, FL, pp. 934–935.

Packard, A., K. Zhou, P. Pandey, J. Leonhardsen and G. Balas (1991). Optimal I/O similarity scaling for full-information and state-feedback problems. *Systems and Control Letters*, **19**, 271–280.

Poolla, K. (1991). Personal communication.

Popov, V. M. (1962). Absolute stability of nonlinear systems of automatic control. *Automat. Remote Control*, **22**, 857–875.

Power, S. C. (1982). *Hankel operators on Hilbert space*. Pitman, London.

Redheffer, R. (1989). Inequalities for a matrix Riccati equation. *J. of Mathematics and Mechanics*, **8**.

Redheffer, R. (1960). On a certain linear fractional transformation. *J. Math. Phys.*, **39**, 269–286.

Rohn, J. and S. Poljak (1992). Radius of nonsingularity. *Mathematics of Control, Signals and Systems* to appear, 1993.

Rotea, M. A., M. Corless, D. Da and I. Petersen (1991). Systems with structured uncertainty: relations between quadratic and robust stability. *Allerton Conference on Communication Control, and Computing*, pp. 885–894; *IEEE Trans. Aut. Control*, to appear.

Rotea, M. A. and P. P. Khargonekar (1989). Stabilization of uncertain systems with norm bounded uncertainty—A control Lyapunov function approach. *SIAM J. Control Optim.*, **27**, 1462–1476.

Safonov, M. G. (1978). Tight bounds on the response of multivariable systems with component uncertainty. *Allerton Conference on Communication, Control and Computing*, pp. 451–460.

Safonov, M. G. (1980). *Stability and Robustness of Multivariable Feedback Systems*. MIT Press, Boston.

Safonov, M. G. (1982). Stability margins of diagonally perturbed multivariable feedback systems. *Proceedings of the IEE*, **129**, 251–256.

Safonov, M. G. (1984). Stability of interconnected systems having slope bounded nonlinearities. *6th Int. Conf. on Analysis and Optimization of Systems*, Nice, France.

Safonov, M. G., (1986). Optimal diagonal scaling for infinity norm optimization. *Systems and Control Letters*, **7**, 257–260.

Safonov, M. G. and J. C. Doyle (1984). Minimizing conservativeness of robustness singular values. In S. G. Tzafestas (Ed.), *Multivariable Control*. Reidel, New York.

Safonov, M. G. and V. X. Le (1988). An alternative solution to the $\mathcal{H}_\infty$ optimal control problem. *Systems and Control Letters*, **10**, 155–158.

Sezginer, R. and M. Overton (1990). The largest singular value of $e^X A e^{-X}$ is convex on convex sets of commuting matrices. *IEEE Trans. on Aut. Control*, **35**, 229–230.

Sideris, A. and R. R. E. de Gaston (1986). Multivariable stability margin calculation with uncertain correlated parameters. *IEEE Conf. on Decision and Control*, pp. 766–773.

Sideris A. and R. Sanchez Peña (1990). Robustness margin calculations with dynamic and real parametric uncertainty. *IEEE Trans. Aut. Control*, **35**, 970–974.

Sideris, A. and R. S. Sánchez Peña (1989). Fast computation of the multivariable stability margin for real interrelated uncertain parameters. *IEEE Trans. Aut. Control*, **34**, 1272–1276.

Skogestad, S. (1987) Personal communication.

Skogestad, S., M. Morari and J. Doyle (1988). Robust control of ill-conditioned plants: high-purity distillation. *IEEE Trans. Aut. Control*, **33**, 1092–1105.

Smith, R. and J. Doyle (1992). Model validation: A connection between robust control and identification. *IEEE Trans. Aut. Control*, **37**, 942–952.

Stein, G. and J. Doyle (1991). Beyond singular values and loopshapes. *AIAA J. of Guidance and Control*, **14**, 5–16.

Wang, W., J. Doyle, C. Beck and K. Glover (1991). Model

reduction of LFT systems. *Proc. of the 30th IEEE Conf. on Decision and Control*, pp. 1233–1238.
Willems, J. C. (1971a). Least squares stationary optimal control and the algebraic Riccati equation. *IEEE Trans. Aut. Control*, **16**, 621–634.
Willems, J. C. (1971b). *The Analysis of Feedback Systems*. MIT Press, Cambridge, MA.
Willems, J. L. (1973). The circle criterion and quadratic Lyapunov functions for stability analysis. *IEEE Trans. Aut. Control*, **18**, 184.
Young, P. (1992) Personal communication.
Young, P. M. and J. C. Doyle (1990). Computation of μ with real and complex uncertainties. *Proc. of the 29th IEEE Conf on Decision and Control*, pp. 1230–1235.
Young, P. M. and J. C. Doyle (1993). Properties of the mixed μ problem and its bounds. In preparation.
Young, P. M., M. P. Newlin and J. C. Doyle (1991). μ analysis with real parametric uncertainty. *Proc. of the 30th IEEE Conf. on Decision and Control*, pp. 1251–1256.
Young, P. M., M. P. Newlin and J. C. Doyle (1992). Practical computation of the mixed μ problem. *American Control Conference*, to appear.
Zames, G. (1966). On the input–output stability of nonlinear time-varying feedback systems, parts I and II. *IEEE Trans. Aut. Control*, **11**, 228–238, 465–476.
Zhou, K. and P. Khargonekar (1987). Stability robustness bounds for linear state-space models with structured uncertainty. *IEEE Trans. Aut. Control*, **31**, 621–623.

a, Vol. 29, No. 1, pp. 111–127, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00
Pergamon Press Ltd

A Fixed H^∞ Controller for a Supermaneuverable Fighter Performing the Herbst Maneuver*†

R. Y. CHIANG,‡ M. G. SAFONOV,§ K. HAIGES,|| K. MADDEN¶ and J. TEKAWY**

A non-scheduled H$^\infty$ *robust flight controller has been designed for a supermaneuverable fighter to fly the Herbst maneuver. The complete design approach and plant uncertainties are documented with detailed linear robustness analysis and nonlinear six degree-of-freedom simulation.*

Key Words—Robust control; supermaneuverable fighter; robust stability; robust performance; H^∞ control; supermaneuverability; Herbst maneuver.

Abstract—This paper presents an H^∞ flight control system design case study for a supermaneuverable fighter flying the Herbst maneuver. The Herbst maneuver presents an especially challenging flight control problem because of its large ranges of airspeed, angle of attack and angular rates. A fixed H^∞ controller has been developed via the mixed-sensitivity problem formulation for 20 linearized models representing the maneuver. Both linear and nonlinear full model evaluations indicate that this single H^∞ controller together with a fixed LQR inner loop feedback have achieved "robust stability" and "robust performance" for the entire maneuver *without* gain scheduling.

1. INTRODUCTION AND BACKGROUND

FUTURE FIGHTER AIRCRAFT development will place increased performance requirements on the design of the flight control system. Maneuvering envelopes are expanding into flight regions characterized by significantly larger levels of modeling uncertainty than encountered in current flight control designs. This expansion of the flight envelope poses a challenging control problem that requires guaranteeing both robust stability and robust performance in the presence of large plant parameter variations, unmodeled dynamics, and nonlinearities. Traditional control design techniques that ignore the effects of these modeling uncertainties will likely produce designs with poor performance and robustness. Recent advances in modern control theory, most notably the H^∞ synthesis technique (Safonov *et al.*, 1989; Chiang and Safonov (1988), and the references therein) offer the promise of a design technique that can produce both high-performance and robust uncertainty-tolerant controllers for next generation aircraft.

To mature this synthesis methodology, the Air Force Wright Research and Development Center (WRDC) initiated the Robust Control Law Development for Modern Aerospace Vehicle program. In this four year program, models for both a hypervelocity and super-maneuverable vehicle were developed, performance specifications and uncertainty models for the two vehicles were defined (Haiges *et al.*, 1989, 1990). The control laws for both vehicles were documented in Haiges *et al.* (1991).

The focus of this paper is to present the H^∞ flight control design methodology developed for the supermaneuverable vehicle. For completeness, a description of the supermaneuverable vehicle and a summary of the performance specification and uncertainty model development effort is also provided.

In phase one of the aforementioned four year Air Force WRDC program, the F/A-18 Hornet was selected as the basis for the super-maneuverable vehicle due to its agile performance and extensive high angle-of-attack aero-

* Received 7 February 1992; received in final form 16 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised from by Guest Editor P. Dorato.

† Invited paper presented in the session " Case Studies of Robust Control Design". *The 29th IEEE Conference on Decision and Control,* Honolulu, HI, 5–7 December 1990. This paper summarized the work done under the project MAVRIC while authors were at Northrop during 1989–1991. This project is sponsored by Wright–Patterson Air Force Base.

‡ Guidance and Control Section, Jet Propulsion Lab., 4800 Oak Grove Dr., MS 198-326, Pasadena, CA 91109, U.S.A.

§ Northrop consultant, also with Dept. of Electrical Engineering systems, University of Southern California, Los Angeles, CA 90089-0781, U.S.A.

|| Aircraft Div., Northrop Corp., Hawthorne, CA 90250, U.S.A.

¶ Integrated System, Inc., Santa Clara, CA 95054, U.S.A.

** Aerospace Corp., El Segundo, CA 90250, U.S.A.

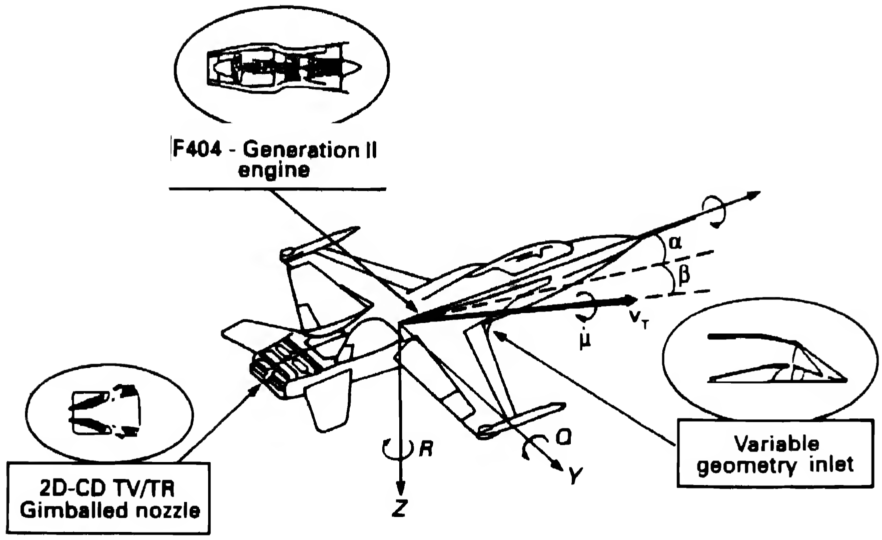


FIG. 1. Supermaneuverable fighter.

dynamic database. A pair of gimballed engine nozzles, which allow in-flight thrust vectoring and thrust reversing were added to the model to enhance its supermaneuverability. This set of additional control effectors enables the vehicle to maneuver with extra axial force and moment control about its three axes (pitch, roll, yaw). Figure 1 shows the basic vehicle configuration.

The Herbst Maneuver (Herbst, 1980) was then selected to demonstrate the H^∞ design technique because of its extreme dynamic behavior: large angles of attack, rapid rotational rate changes, large angular accelerations, and wide variations of dynamic pressure and speed. As described in Herbst (1980) the maneuver can be initiated from level flight at an altitude of 10,000 ft and a velocity of 300 ft sec^{-1}. A pitch-up is commanded until a maximum angle of attack of 80° is reached. This causes flight path angle and altitude to increase and causes velocity to decrease rapidly until the wing stalls. At a post-stall velocity of 50 ft sec^{-1}, and 85° bank angle command is initiated using engine thrust vectoring to rotate the flight path heading by 180°. Once the aircraft begins to roll, the angle of attack is decreased and velocity is increased. At this point, the maneuver is complete. Figure 2 shows the actual maneuver trajectory.

In phase two, the flying quality specifications and uncertainty models were developed. A performance specification for the regulator design was also developed from the maximum allowable dynamics envelope described in Hodgkinson (1982). The performance design constraint is to achieve a feedback regulator with closed-loop singular-value frequency-response bounded by such an envelope.

The sources of uncertainty in this maneuver were divided into four different types:

Type 1. Uncertainty in the aerodynamic data base from wind tunnel test error, which includes Y_β, L'_β, L'_P, L'_R, M_Q, N'_β, N'_P, N'_R, $Y_{\delta R}$, $L'_{\delta A}$, $L'_{\delta R}$, $N'_{\delta A}$, and $N'_{\delta R}$ (see McRuer et al. (1973) for definitions of standard aerodynamic derivatives).

Type 2. Time delay at plant outputs (12.5 msec).

Type 3. Neglected sensor/actuator dynamics (14 actuators, five sensors).

Type 4. Variations in structural mode frequency (±5.5%), damping (±31%), and shapes (±10 ~ 15%) due to mass variations.

Type 5. Variations in mass and inertia properties (±10 ~ 24%).

The robustness design objective is to fly the Herbst maneuver with robust stability in the presence of all these uncertainties. Together with the above performance specifications, the regulator design should provide "robust performance" throughout the maneuver envelope.

In this paper, a novel H^∞ multi-model design concept is presented. It is shown that with a robust equalization inner loop, an eight-state 3×3 mixed-sensitivity H^∞ controller (feeding back angle of attack α, sideslip angle β, and bank angle rate $\dot{\mu}_{rot}$) to appropriate linear combinations of aerodynamic and engine-thrust-vectoring actuators) is capable of flying the Herbst maneuver without gain scheduling. This special "non-scheduling" capability provides immediate design advantages:

- Reduced software/hardware implementation effort.
- Lower software/hardware maintenance requirements.
- Guaranteed robust performance throughout the maneuver envelope.

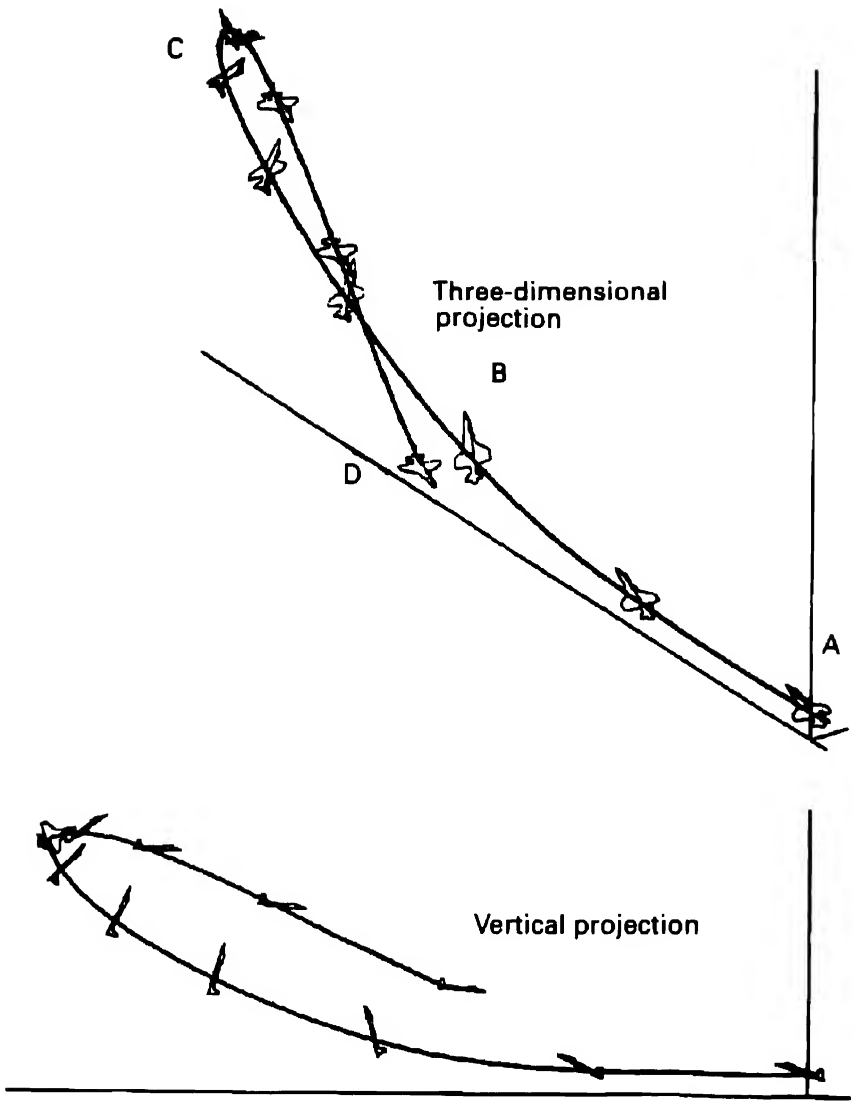


FIG. 2 Herbst-type maneuver

The latter is possible because the time-variations in the controller due to gain-scheduling are eliminated and time variations in the model are greatly reduced by the robust equalization inner-loop. The net effect is that time-variation is reduced to the point that conditions of small-gain stability robustness and performance-robustness criteria can be made to hold for all time-variations in the plant using a linear-time-invariant H^∞ controller.

Section 2 describes the general multi-model flight control design methodology. Section 3 shows the results of applying this method to the design of Herbst maneuver flight control system. Conclusions are presented in Section 4.

2. CONTROL LAW DESIGN APPROACH

The objective of the flight control design is to find an H^∞ robust controller capable of flying the Herbst maneuver with robust performance in the presence of the uncertainties described in Section 1.

2.1. *Control structure*

The overall structure of the flight control system uses an explicit model-following approach inherited from the DMICS (Design Methods for Integrated Control Systems) concept presented in Shaw *et al.* (1988). Figure 3 shows the block diagram, which contains the following functional elements:

(1) Maneuver Command Generator (MCG)—$M(s)$.

(2) Configuration Management Generator—CMG.

(3) Feedforward controller—$K_2(s)$.

(4) Feedback regulator—$K_1(s)$.

(5) Control selector—CS.

The maneuver command generator (MCG) is the forward path model of the explicit model-following control system. Its output signals are used to form the error and feedforward terms for the model-following regulator. The MCG is designed to directly embody the desired flying quality characteristics

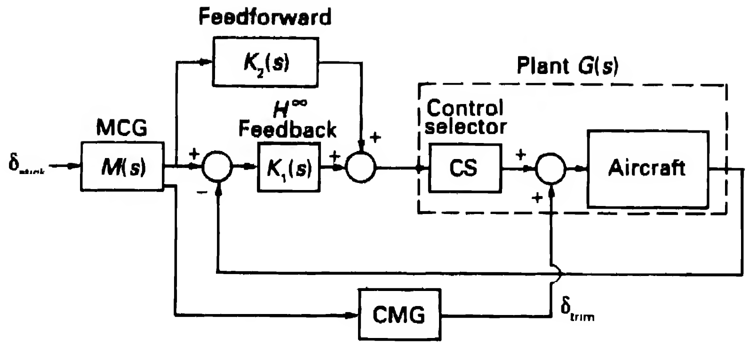


FIG. 3. General flight control block diagram.

for each mode of flight. The primary inputs for the MCG are the pilot's controller commands (e.g. longitudinal stick force). Its outputs are open-loop time-histories of various aircraft attitude angles and rates which, if perfectly tracked by the H^∞ feedback controller $K_1(s)$, result in an ideal maneuver. For the controller which we designed, the signals output by the MCG were commands for angle of attack (α), sideslip angle (β), and bank angle rate ($\dot{\mu}_{rot}$).

The functions of feedback regulator $K_1(s)$ are to:

(1) precisely track the MCG commands;

(2) reject plant disturbances;

(3) reduce sensitivity to plant parametric variations;

(4) provide robust stability for the sytem.

The first three tasks can be handled by minimizing the sensitivity function $S = (I + GK_1)^{-1}$. In flight control, the fulfillment of fourth requirement can be facilitated by a nonlinear control selector (CS) which accepts as its inputs control generalized controls (which are equivalent to angular accelerations in pitch, roll, and yaw) and generates appropriate commands to the physical control effectors on the aircraft depending of the flight condition (Shaw *et al.*, 1988). The use of generalized controls provides a decoupling between the design and implementation of control law gains from the selection of specific effectors for producing the desired control accelerations. In addition, the regulator design is simplified by inclusion of the CMG (configuration management generator). The CMG is a nonlinear element which generates an approximation to the desired trim solution throughout the flight envelope, so the linear flight control system regulator functions as a perturbational device.

The feedforward controller K_2 is a command response shaping filter which can be used to speed up the aircraft response if the regulator is too sluggish. If the regulator has a faster control bandwidth than the MCG, K_2 can be dispensed with, without degrading the command response. In fact, the Bode sensitivity theorem indicates that most of the feedback issues such as disturbance rejection, plant parametric sensitivity reduction and so forth are not a function of $K_2(s)$, rather, they are only a function of the feedback regulator $K_1(s)$.

The control selector (CS) is used to distribute to the aircraft's control effectors the generalized control moment commands ($\dot{P}$, $\dot{Q}$, $\dot{R}$)—P, Q, R are standard notation for the rates of change of and aircraft's roll, pitch and yaw angle, respectively (McRuer *et al.*, 1973). The control selector attempts to neutralize the effects of changes in dynamic pressure or engine thrust so that the effective control forces and moments are constant throughout the maneuver. It is a nondynamical nonlinear device designed to compensate for some of the aircraft's nonlinearities so that the overall response of the aircraft angular rates P, Q, R to inputs to the control selector is very nearly $\frac{1}{s} I_{3\times 3}$ in the mid-frequency range, below wing and fuselage bending mode frequencies and above the frequency of aerodynamic response modes—about 3–20 rad sec^{-1} in our case. Therefore, its gains are a function of the vehicle's operating state (e.g. Mach, altitude, thrust setting) and are scheduled throughout the maneuver. It also can provide control redistribution in the event of actuator saturation or failure (Shaw *et al.*, 1988). In general, the control selector transformation is based upon available control power.

The main issues in the control design—command issues and feedback issues—are addressed by the MCG, the feedforward controller, and the feedback regulator. It was proved in Pernebo (1981) that, for linear control systems, the command issues can be separated from feedback issues completely. In other words, the nominal command response can be "shaped" by a prefilter and feedforward (in the present case, also collectively known as the

MCG) both designed after the feedback controller loops. The particular choice of feedback in no way limits or changes the set of nominal command response transfer functions that are attainable via feedforwards and/or prefilters. This philosophy was executed throughout the design process. In this paper, we focus primarily on the "feedback issues" of the problem, i.e. how to design a feedback regulator with good tracking performance and large stability robustness.

2.2. Regulator design—performance and robustness

Regulator design has two fundamental objectives, Robust Tracking Performance and Robust Stability. To achieve robust performance, one needs to define a proper feedback structure and a set of suitable performance specifications. To accomplish robust stability, one has to determine a set of linear models to be stabilized in the presence of uncertainties in pre-defined sets. The following sections elaborate these concepts in detail.

Performance specification. As discussed in Section 2.1, the MCG is designed to embody desirable flying qualities properties for each mode or condition of flight. This is done by utilizing command filters patterned after the low-order equivalent system (LOES) transfer functions for aircraft response that are described in the military standards (MIL-STD-1797, 1987).

In Hodgkinson (1982), the author defines frequency-domain limits on the maximum allowable dynamics that may be added to the ideal LOES transfer functions without degrading the flying qualities. This envelope of maximum allowable dynamics, $N(s)$, is shown in Fig. 4. For the study described herein, $N(s)$ has been interpreted as a performance specification as follows.

Let $M(s)$ be the desired "ideal" aircraft response, e.g. as described by the military standard LOES transfer functions

$$y = M(s)\delta_{\text{stick}},$$

where y is the vector of output to be controlled and δ_{stick} is the pilot's control stick position. By definition, when the maximum allowable unnoticeable dynamics are added to the system this relation becomes

$$y = N(s)M(s)\delta_{\text{stick}}.$$

Since the MCG is designed to provide the desired "ideal" flying qualities, the transfer function of the MCG may be substituted for $M(s)$.

From the block diagram in Fig. 3, it is found that the overall input to output relation for the system is given by:

$$y = N_1(s)M(s)\delta_{\text{stick}},$$

where

$$N_1(s) = (I + GK_1)^{-1}G[K_2 + K_1].$$

Thus, it is seen that if $N_1(s)$ can be maintained within the envelope of $N(s)$, the added dynamics from the closed-loop tracking system will be unnoticeable to the pilot. Hence, the gain/phase envelope of $N(s)$ becomes the "feedback performance specification" for the regulator design.

As a final note, the $N(s)$ specification described in Hodgkinson (1982) was developed for aircraft response in the pitch axis. However, due to a lack of available standards for other performance specifications, it was adopted as the specification for all of the regulator channels in our design. Proper bandwidth scaling between position and rate variables will be imposed at a later stage of the H^∞ design.

In supermaneuvering flight, the rate of roll

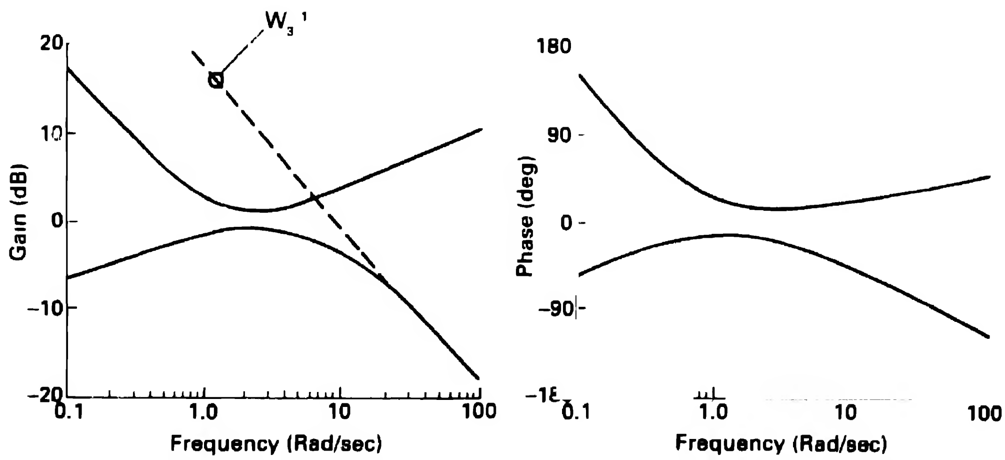


FIG. 4. Specification of maximum unnoticeable dynamics and H^∞ weighting strategy.

$\dot{\mu}_{rot}$ of the aircraft body about its velocity vector V_T is important to be fed back for tracking performance—see Fig. 1. In particular, the so-called "rotational part" of this variable defined by

$$:P\cos(\alpha) + R\sin(\alpha), \qquad (1)$$

provides a useful measure of what a pilot would like to control during high angle-of-attack rolling such as in the Herbst maneuver. Basically, in such large angle maneuvers, neither the body x-axis nor the velocity vector (wind x-axis) can be used as the roll-axis reference, they change directions themselves. Thus, to achieve a coordinated maneuver, controlling body roll-axis motion P has been replaced by control of the rotation about the total velocity vector. The variable $\dot{\mu}_{rot}$ described in (1) representing a combined motion of roll and yaw (with zero sideslip angle) serves this purpose satisfactorily. In other words, precisely controlling this variable $\dot{\mu}_{rot}$ enables the aircraft to perform a "coordinated" large angle roll maneuver with zero sideslip angle. Details of this concept can be found in Kalviste (1986). This "decoupled" response (roll with zero sideslip angle) is one of the fundamental requirements of tracking performance.

Uncertainty models. As described in Section 1, five sources of uncertainty were identified for the current design study. The aerodynamic uncertainties were based upon the differences between two aerodynamic data bases—wind tunnel and flight test data. This formed the structured uncertainty models for the aircraft stability and control derivatives. For the purposes of H^∞ controller design, the types 2–5 uncertainties were treated as unstructured. However, in the subsequent linear evaluation of our design, it proved convenient to exploit the fact that in actuality the type 4 and 5 uncertainty in our problem turns out to be dominated by a known nonlinear function of a single real parameter, viz. the total weight of the aircraft (which varies as fuel is consumed).

Type 2 uncertainty was treated as unstructured multiplicative uncertainty at the plant outputs. The neglected structural modes and sensor dynamics were considered as frequency bounded multiplicative uncertainty at the plant outputs.

Linear model development. Twenty 56-state linear-time-invariant models were generated by linearizing a faithful high-order nonlinear simulation model at various points along the nominal Herbst-type maneuver trajectory described in Section 1. As shown in Fig. 2, position A is the starting point. Position B is the design point due to its "central position with respect to all the other linear models (this will be elaborated in Section 2.3). Position C is the turning point, which is also considered to be the "worst case" due to its rapidly changing aerodynamics. At position D, the maneuver is complete. All the aerodynamic data were based on a full F/A-18 database that was integrated into the nonlinear simulation model. The load condition of the aircraft was characterized as carrier landing gross weight with empty fuel tank.

The 56 states in the linear models from position A to D represent actuators, sensors, structural modes, engine states, and the standard six degrees-of-freedom flight dynamics. The available control actuators are stabilities, ailerons, rudders, pitch and yaw thrust vectoring. Throughout much of the maneuver, the engine thrust vectoring actuators provide most of the control energy, since other regular control surfaces are not as effective at the relatively low airspeeds which are characteristic of the post-stall Herbst-type maneuver. The distribution of control energy to the available control actuators has been carefully designed via the control selector described in Section 2.1.

The standard offset derivative technique was used to linearize the aircraft model about each particular operating point. Since the operating point continuously changes throughout the maneuver, one obtains a different linearized model for each time. Thus, although each such model is a linear-time-invariant model, the model varies with time. Furthermore, the rate of time variation is only slightly slower than our design bandwidth. In other words, the models might not be adequate to predict the stability and performance at each operating point along the nonlinear maneuver trajectory. Fortunately, after the robust equalization inner-loop feedback is closed, the amount of variation between models becomes sufficiently small that the small-gain stability theorem can be used to guarantee both stability and performance robustness throughout the maneuver.

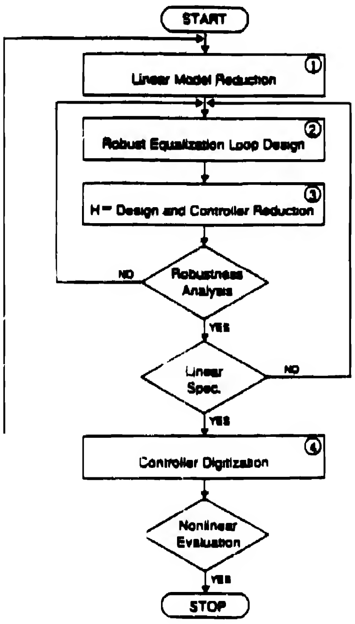
Model reduction is always needed to get a suitable set of states to be stabilized by the regulator. This will be elaborated in Section 3 for the Herbst design.

2.3. *Multi-model design concept*

After the linear models and design requirements are developed, the actual design process can be shown in Table 1.

In this section, Steps 2 and 3 are discussed in detail. Results for a continuous controller that demonstrates only the major design concept are presented in this paper. The digitization effects involved in Step 4 will be examined later. Model

TABLE 1. DESIGN FLOW CHART



reduction for this design is presented in Section 3. Further discussion on the model reduction and its robustness issues can be found in Safonov and Chiang (1988b).

Robust equalization loop design. This particular step plays a crucial role in the entire design process. If the H^∞ design procedure had been directly applied to various linear models along the trajectory individually, each H^∞ design could have only assured robust performance for one single linear model. Moving one controller to another operating point could not be guaranteed to produce a stable response. This is because other models along the trajectory are never considered in the single point design process. Implementation in the traditional fashion would require massive gain scheduling of the controller throughout the maneuver. Neither the system robustness could be guaranteed after the gain scheduling, nor would a simple fixed linear-time-invariant controller implementation be possible.

A new multi-model design concept that employs an extra inner loop can remove the disadvantages and inadequacies of the gain-scheduled "single" point design approach. It has long been known that feedback can reduce the sensitivity to plant uncertainty. Zames (1981) characterized this property in terms of shrinking an H^∞ ball of uncertainty with feedback. With this concept in mind, the role of inner loop becomes obvious: It can "shrink" the "distance" between different models, thereby reducing the modeling uncertainty between them.

Our design process began with analyzing the "distance" between linear time-invariant models of the plant $G(s)$ obtained by linearizing at several different along the nominal Herbst maneuver trajectory. Recall that the control selector has been designed so that the mid-frequency range ($3\,\text{rad sec}^{-1} < \omega < 20\,\text{rad sec}^{-1}$) response to P, Q, R is approximately $(1/s)I_{3\times3}$. With the control selector in place, this same mid-frequency response in the P, Q, R channels occurs irrespective of the aircraft attitude and velocity throughout the entire maneuver. Of course, at lower frequencies the responses may vary substantially. However, the fact that the mid-frequency responses are all about the same greatly eases the robust control design task. It means that a fixed pure-gain inner-loop controller having loop singular value Bode plot magnitude crossover frequencies in this mid-frequency range has the potential to stabilize all the plants, in addition to drawing their low-frequency-range frequency responses more closely together, shrinking the distance between them so that they are all fit inside a smaller H^∞ ball. Because we have that all the plant variations have about the same $1/s$ mid-frequency-range behavior and because the low-frequency differences though drastic, are not so extreme as to alter the phase by more than about 90° in any channel, a pure gain or simple LQR inner-loop feedback with crossover in the 3–$20\,\text{rad sec}^{-1}$ range can be expected to do this robustly. We decided to use an LQR inner loop feeding back the five states α, β, P, Q, R of the nominal reduced-order linearized plant G. The LQR penalty function used was $\int_0^\infty (x'x + \rho u'u)\,dt$ (with the scalar $\rho > 0$ chosen using the LQR return-difference equality as a guide) to produce an inner-loop crossover frequency of about $10\,\text{rad sec}^{-1}$. Examining the response to the α, β, $\dot{\mu}_{\text{rot}}$ channels, we found, not unsurprisingly, that the $10\,\text{rad sec}^{-1}$ inner-loop also shrinks the distance between these models.

To measure the distance between two models, say G and G_0, we used the spectral radius of the multiplicative error

$$|\lambda_{\max}(\Delta(j\omega)|,$$

where $\Delta = (G - G_0)G_0^{-1}$. However, if the two plants have different numbers of right-half plane poles the distance between them is defined to be infinite.

The following inequality gives some indication why the multiplicative error, instead of the absolute error $|\lambda_{\max}(G - G_0)|$, is chosen for the inner loop design:

$$\begin{aligned} |\lambda_{\max}[\Delta]| &= |\lambda_{\max}[(G - G_0)G_0^{-1}]| \\ &\le |\lambda_{\max}[G - G_0]|\,|\lambda_{\max}[G_0^{-1}]|. \end{aligned}$$

Clearly, the multiplicative error can be large even when the absolute error is small if

$\lambda_{max}[G_0^{-1}]$ is large, as is usually the case at high frequencies.

Why is the multiplicative error important in multi-model design? The reason is that when the multiplicative error between models is small, it guarantees that the logarithmic Bode plot of a perturbed model stays inside an error band along the nominal model, which from the control design point of view is a way of assuring robust performance for both models with single feedback control law. In papers by Safonov and Chiang (1988a) and Safonov et al. (1990) the role and use of the multiplicative error was elaborated in detail. In particular, a "Robustness Theorem" presented in these papers provided a sufficient condition for several models to be stabilized by a single feedback. (Another approach, involving the "Gap Metric", which is mathematically much more subtle but accommodates uncertain right-half plane poles, might also have been used to measure the distance between two models (Georgiou and Smith, 1990). However, as all our plant models turned out to be stable, gap metric analysis was unnecessary.)

High performance H^∞ *loop design*. The standard mixed-sensitivity weighting strategy (Safonov and Chiang, 1988a) was used in our H^∞ problem formulation

$$\min_{F(s)} \left\| \begin{matrix} W_1 S \\ W_3 T \end{matrix} \right\|_\infty \le 1, \tag{1}$$

where the low-frequency loop transfer function ($L(s) = GF$) is shaped by penalizing the sensitivity function $S = (I + GF)^{-1}$ to reject the plant disturbances and noises and the high-frequency $L(s)$ is shaped by penalizing the complementary sensitivity $T = GF(I + GF)^{-1}$ to cope with uncertain actuator nonlinearities, unmodeled structural modes and sensor/actuator dynamics. Figure 5 shows the plant augmentation strategy associated with this approach. The

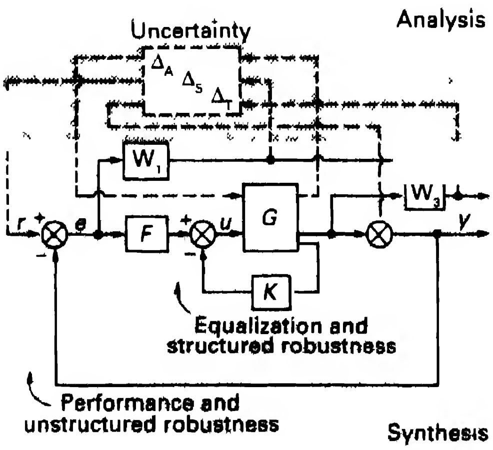


FIG 5 H^∞ synthesis and analysis block diagram

numerically robust loop-shifting formulae developed in Safonov et al. (1989) were used to find the optimal H^∞ controller.

A salient feature of the mixed sensitivity approach is that it treats the uncertainty as unstructured. The frequency dependent weights $W_1(s)$ and $W_3(s)$ are used by the control designer to specify the desired loop-shapes. The "all-pass" property of optimal H^∞ control laws ensures that the singular value Bode plots will precisely conform to those specified by $W_1(s)$ and $W_3(s)$ (see Safonov and Chiang (1988a) for details). However, there is a serous drawback to this property, namely, the resulting mixed-sensitivity H^∞ controller always blindly cancels the LHP poles of the plant, which can seriously damage the system robustness against structured (parametric) uncertainties associated with uncertain lightly-damped pole locations, even though the overall loop transfer function is shaped exactly to the specifications (weighting functions W_1^{-1} and W_3^{-1}). As for the reason why there exists such an unpleasant phenomenon, Chiang et al. (1990) examined the SISO case and Sefton and Glover (1990) examined the MIMO case.

The robust equalization loop removed this difficulty completely. It provided large multivariable stability margin against structured parametric uncertainty of the plant by damping the uncertain lightly-damped "dutch-roll" modes with inner loops. This automatically resolves the problem of lightly-damped pole/zero cancellation associated with the mixed-sensitivity H^∞ problem formulation. The result is shown in Section 3. This inner-loop strategy for increasing the plant parametric robustness was first proposed in Chiang et al. (1990). Compared to other approaches such as μ-synthesis (Doyle et al., 1987) it is our opinion that this approach offered a simpler and more direct solution in this application.

Exact "feedback decoupling" is a non-trivial design task (Falb and Wolovich, 1967; Isidori et al., 1981). However, via the mixed-sensitivity H^∞ design method described above, feedback decoupling becomes an extremely simple task. Inside the control bandwidth $|W_1(j\omega)| \gg 1$ and the constraint on the cost (1) imposed by $W_1(s)$ ensures that the sensitivity function ($S = (I + GF)^{-1}$) will be smaller than $|/1W_1| \ll 1$, thus ensuring that the closed-loop transfer function $T = I - S$ will be approximately equal to the identity matrix (I). Clearly, I is diagonal and therefore decoupled.

3 HERBST MANEUVER DESIGN

An H^∞ controller has the same order as the plant, so it is desirable either to approximate the

H^∞ controller by a reduced order controller as final design step or to use a reduced order plant model so that the resultant H^∞ controller will be of correspondingly reduced order. We chose the latter approach.

The 56-state full order model was reduced at design stage by applying the following simplifications.

(1) Sensors and structural modes are fast as compared to design bandwidth ($10\,\text{rad sec}^{-1}$), and may therefore be neglected.

(2) Actuators dynamics and nonlinear effects can be neglected in the plant models and accounted for in the uncertainty specification.

(3) The engine states are decoupled from the other dynamics in this maneuver, and may be removed.

(4) The trajectory related states such as Euler angles (ϕ, ψ, θ), altitude (h), and total velocity (V_{tot}) naturally diverge during maneuvering flight and therefore should not be stabilized by feedback. Instead one should stabilize the rotational states: angle of attack (α), sideslip (β), and angular rates (P, Q, R) so that the pilot/MCG may guide the aircraft through the desired maneuver by commanding those trajectory states.

A singular perturbation model reduction routine was used to remove the unwanted states, leaving the linear model with five states (α, β, P, Q, R) to be stabilized by the regulator.

Further simplification of the linear models was permitted due to the use of the decentralized DMICS control structure described in Section 2.1. For the regulator design, the control selector is assumed to be "perfect". That is, the demanded control forces (i.e. the generalized controls) are exactly achieved by the aircraft's aerodynamic and propulsive control effectors. With this assumption, the rows of the "B matrix" of the linear model corresponding to the $\dot{P}$, $\dot{Q}$, $\dot{R}$ state derivatives may be replaced by a generalized control effectiveness matrix (a rank three standard basis matrix), which includes three inputs and allows direct command of $\dot{P}$, $\dot{Q}$, and $\dot{R}$.

Since all the states were available in our model, a fixed LQR inner loop gain with unity "state" and "control" weighting was then developed to the design model (position B) and the worst case model (position C) of the maneuver. Using "unity" weighting keeps the control energy normal and leaves the system bandwidth unchanged. If the unity LQR weights are not enough to minimize the multiplicative error between models, a heavier weighting strategy can be employed until the goal—multiplicative error is less than one, is achieved.

Figure 6 shows that the multiplicative error between the five-state "design model" (at maneuver point B) and the 51-state "evaluation models" (at maneuver points A, C and D without five trajectory states) has been reduced to less than unity throughout the control bandwidth 0–$10\,\text{rad sec}^{-1}$ after the unity weighting LQR inner loop is closed. In other words, all the linear models have been approximately equalized by a single LQR feedback. Moreover, this fixed LQR gain also stabilized all the linear models (either five-state or full 56-state) along the maneuver from point A to D (see Fig. 7). The robustness theorem of Safonov and Chiang (1988b) implies, roughly speaking, that a controller will produce similar performance for all plants whose deviation from the nominal has a multiplicative error of less than about -3 dB inside the control loop bandwidth. We see from Fig. 6 that, with our robust equalization inner-loop in place, this conditions holds for controllers having bandwidth up to about $10\,\text{rad sec}^{-1}$. Consequently, it follows from the robustness theorem that a fixed H^∞ controller should be able to accomplish the outer loop design without gain scheduling.

It is well known that linear quadratic regulator feedback is highly robust with guaranteed $(0.5, \infty)$ gain margin and ± 60 degree phase margin in every feedback channel (Safonov and

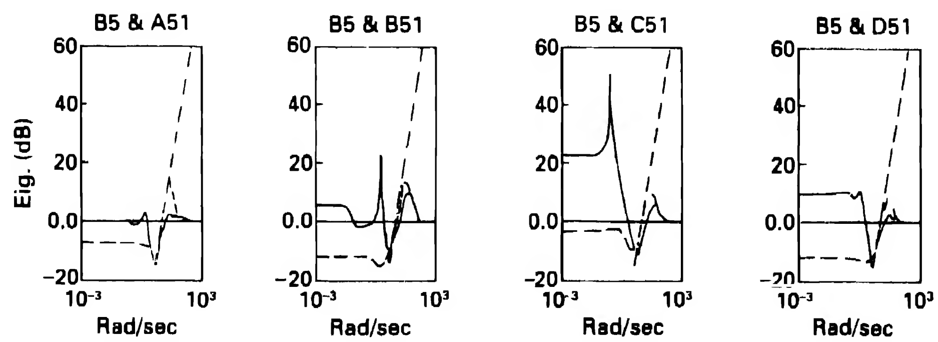


FIG. 6. Multiplicative error—five-state model B vs full (51)-state models A, C, D (solid: without LQR; dashed: with LQR; dash-dot: W_1 weighting).

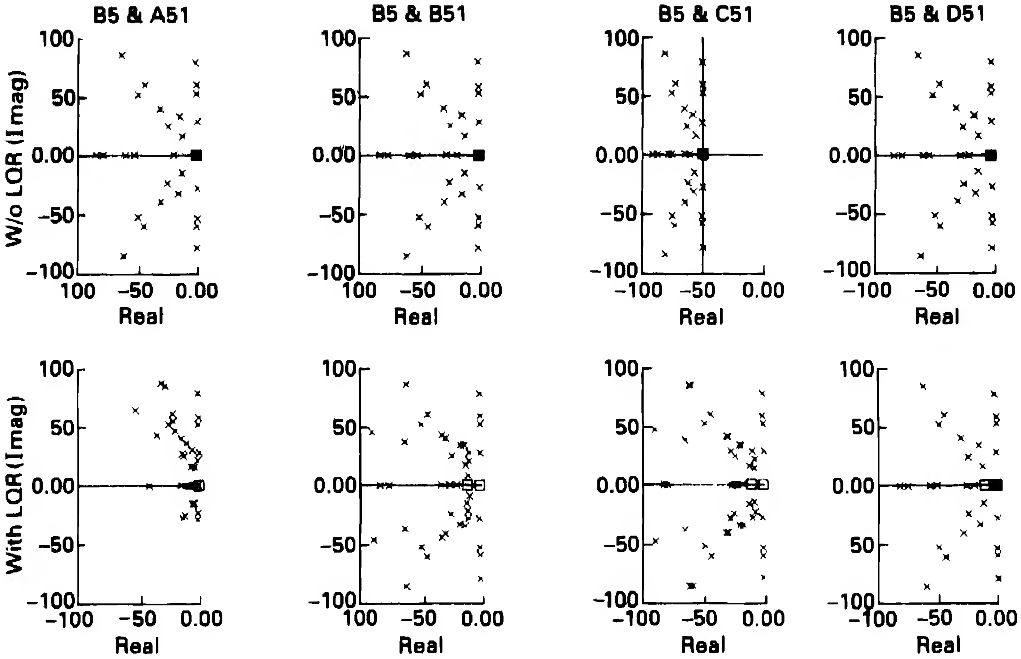


FIG. 7. Pole locations with and without LQR (×; G_0, □: G).

Athans, 1977). In our case, a fixed LQR gain has created an extra benefit for the outer loop design, namely, it makes the outer loop H^∞ design much more robust against plant parameteric uncertainty.

The H^∞ weighting matrix W_3 was chosen as $W_3 = \mathrm{diag}\,(100/s^2, 100/s^2, 10/s)^{-1}$, which roughly coincides for $\omega > 10$ with the lower bound from the specification constraint on maximum unnoticeable dynamics shown in Fig. 4. Although a higher bandwidth also meets the constraint, the aircraft control design does have high-frequency constraints at and beyond $10\,\mathrm{rad\,sec^{-1}}$ (due to actuator rate limits, unmodeled structural modes, etc.) that are addressed by the high-frequency roll-off imposed by W_3. In other words, the choice of W_3 limits the H^∞ design bandwidth to about $10\,\mathrm{rad\,sec^{-1}}$ so that the high-frequency unstructured uncertainties cannot drive the system into instability (see Fig. 6). The different roll-off rates were used to ensure that the positions (α, β) and rate ($\dot{\mu}_{rot}$) channels are penalized equally. Without such consideration, high-frequency lead would enter the control dynamics, amplifying sensor noise and causing undue actuator saturation.

The low-frequency penalty W_1 was chosen to be

$$W_1 = \rho\left(\frac{s + 0.01}{0.02s + 1} I_3\right)^{-}$$

where the DC gain ρ is used as a design "knob" to push the sensitivity function S down as much as possible, until the complementary sensitivity T is pushed against W_3^{-1}. Since the H^∞ controller will contain the dynamics of W_1^{-1}, the zero location (0.01) was chosen for the low-frequency range disturbance rejection, the pole location (100) was chosen to ensure a well-posed state-space solution of the H^∞ mixed-sensitivity problem (Chiang and Safonov, 1988). As a consequence of minimizing the sensitivity function, commands will be tracked well and wind gust disturbances can be substantially attenuated.

Some output scaling between position variables and rate variables is necessary in multivariable control system design. All of the rate variables (P, Q, R, $\dot{\mu}_{rot}$) are scaled down by a factor of ten, so that ten degrees-of-position deflection are comparable to one $\mathrm{deg\,sec^{-1}}$ rate increment (since the design bandwidth is approximately $10\,\mathrm{rad\,sec^{-1}}$).

3.1. *Design results*

Linear evaluation. In this section, the results of nominal design (five-state at position B) and the results of connecting the same controller with the evaluation models (51-state at positions A, C, D) are shown.

Figure 8 illustrates the "perfect" loop shaping S/T functions and step response for the design model (five-state at point B). Figure 9 shows the loop shaping for different operating points. Notice that the shapes of the sensitivity functions indicate that the wind gust disturbance at frequencies below $1\,\mathrm{rad\,sec^{-1}}$ will be attenuated by a factor of 1000 throughout the maneuver. Figure 10 shows the step responses. Clearly, the

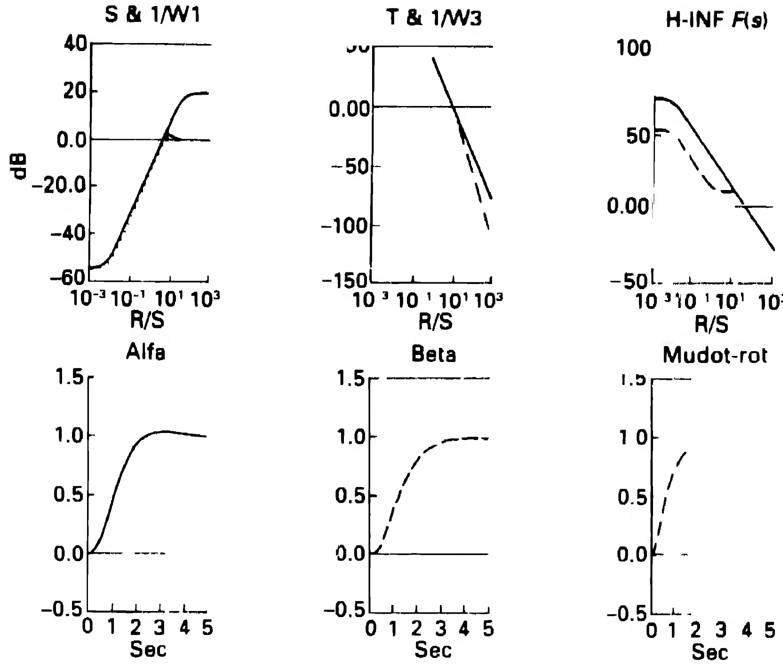


FIG 8 Results of H^{∞} design for five-state nominal model at position B

tracking loops (α, β, $\dot{\mu}_{rot}$) are completely decoupled at design point (position B) and are coupled around the worst case operating point (position C).

With the robust equalization inner loop, it is not surprising to see that at a particular operating point, the design can tolerate a lot of structured parametric uncertainty, yet still can maintain the outer loop tracking and reject the high-frequency unstructured uncertainties.

Numerically reliable algorithms (Tekawy *et al.*, 1989) were used to compute the Structured Singular Value (SSV) of the multivariable stability margin seen by the aerodynamic uncertainties. The 51-state evaluation models were used here as well. Figure 11 shows the complete interconnection of robustness test. The dashed lines in Fig. 12 show that in such a high angle-of-attack and low dynamic pressure maneuver, the system with a "robust equalization inner-loop" has "robust stability" against all the anticipated uncertainties of type 1, 2 and 3. Uncertainty types 4 and 5 turn out to depend almost entirely on a single parameter, viz. the aircraft total weight (as determined by its current fuel load). Accordingly, to evaluate the effects of including these additional uncertainties, it sufficies to examine how the SSV plots for uncertainty types 1–3 vary as aircraft weight varies. We did this for several possible values of aircraft weight and found that the SSV plots varied little and remain below 0 dB as the weight varied between its minimum and maximum values as shown in Fig. 13. Thus robust stability is guaranteed for all of the uncertainty types 1–5.

The set-up for "robust performance" evaluation is shown in Fig. 11, where the H^{∞} weighting filter W_1 is cascaded with an extra *fictitious* uncertainty block (Δ_{perf}) and evaluated simultaneously with the rest of uncertainties ($\Delta_{A,B}$, Δ_{ACT}, Δ_{SEN}) (Doyle *et al.*, 1982). The solid lines in Fig. 12 are the results of robust performance evaluation for 51-state models at position A, B, and C of the maneuver with the aircraft at its nominal weight—similar plots result for other values of total aircraft weight between its minimum and maximum (type 4 and 5 uncertainty). This implies that this design can track commands accurately in the presence of all the anticipated uncertainties (types 1–5).

(Plant and uncertainty data are available from the authors. Interested parties may recieve the data by sending a blank 3.5 in 1.44 MB IBM floppy and a self-addressed stamped envelope.)

Nonlinear evaluation. After the linear evaluation of the design, it is necessary to check the robust performance from a nonlinear perspective. The main objective at this stage of the design was to evaluate the robust performance of the H^{∞} design in the presence of the nonlinear aerodynamics of the Herbst maneuver. Other issues including actuator rate limits, maneuverability and kinematic coupling effects can also be explored.

The eight-state optimal H^{∞} controller was implemented in our complete nonlinear simula-

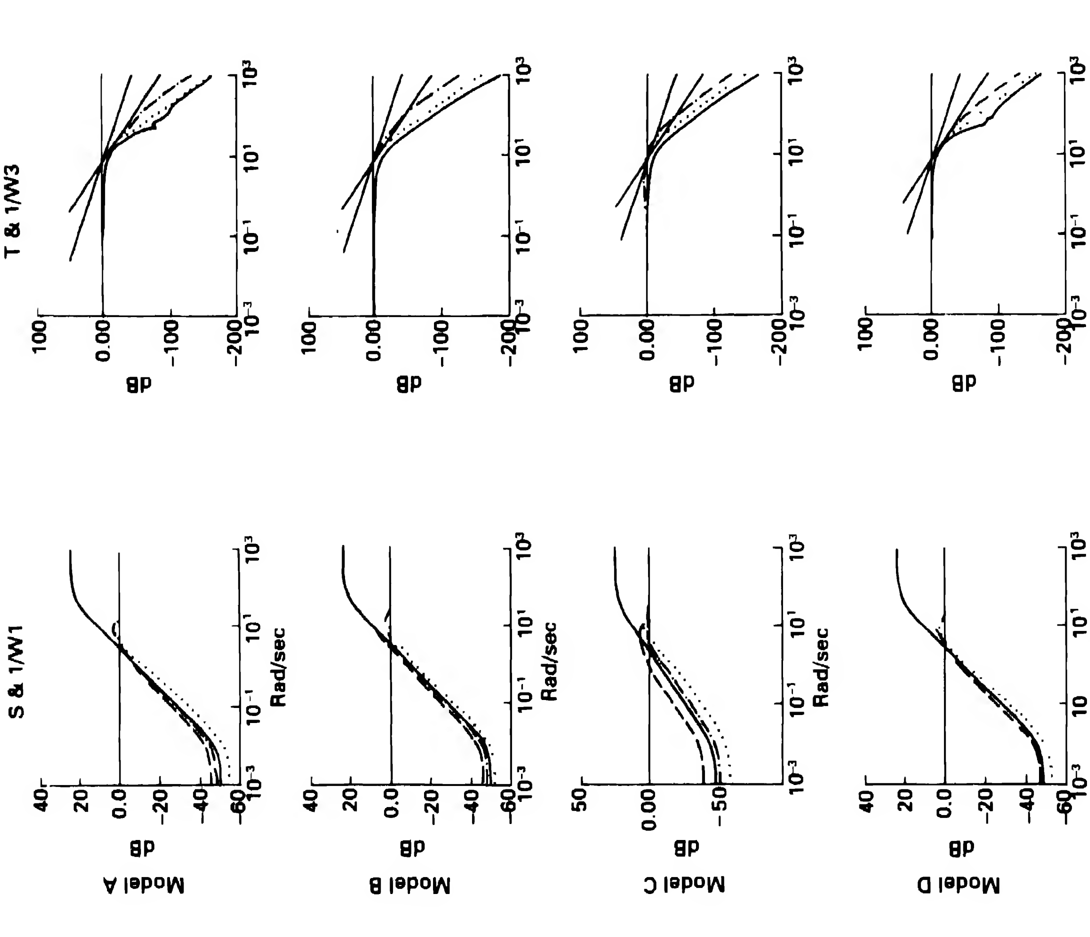


FIG. 9. Results of S and T for 51-state evaluation models at position A, B, C and D.

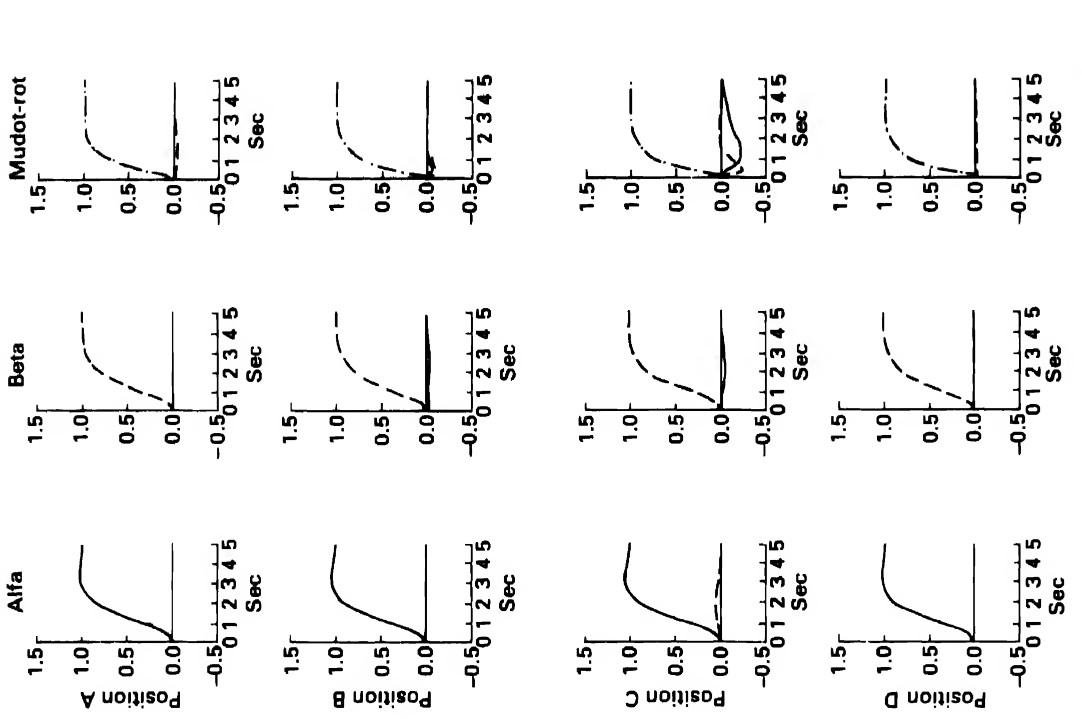


FIG. 10. Step responses of 51-state models in α, β and $\dot{\mu}_{rot}$ command channels.

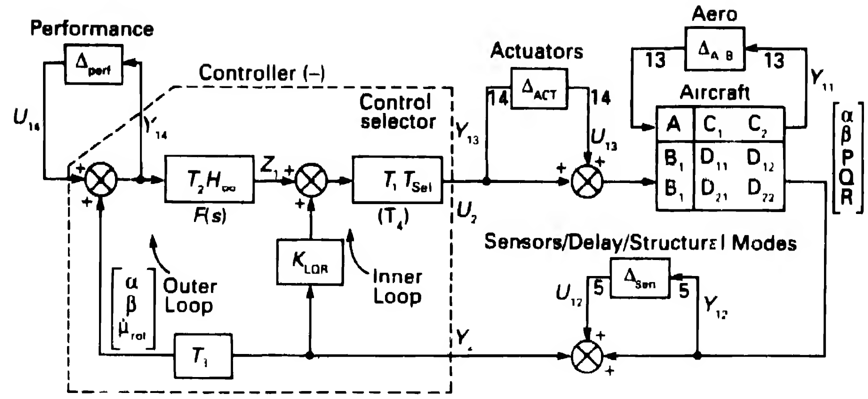


FIG 11 Block diagram for robustness evaluation

tion program with the standard six degrees-of-freedom equations of motion, nonlinear actuators, sensors and structural modes. The MCG and CMG were also implemented to generate the proper command profile of the Herbst maneuver.

In Fig. 14, an angle-of-attack command and a roll pulse command were generated from the MCG. The engine level was set to full-throttle to build up the proper required angle of attack for this nameuver. Figure 15 shows the associated angular rates of the airplane. As shown in Fig. 16, once the angle of attack reached its maximum (78°) and the total velocity dropped down to its minimum ($55\,\mathrm{ft\,sec^{-1}}$), a roll pulse command ($\dot{\mu}_{rot}$) was initiated to rotate and turn the aircraft heading 180°. Figure 17 shows the responses of the control surfaces—horizontal tails (δ_H), leading/trailing edge flaps (δ_F, δ_N) and thrust vectoring (δ_{TV}). They are all well within saturation limits as expected.

As the linear analysis predicts, the following

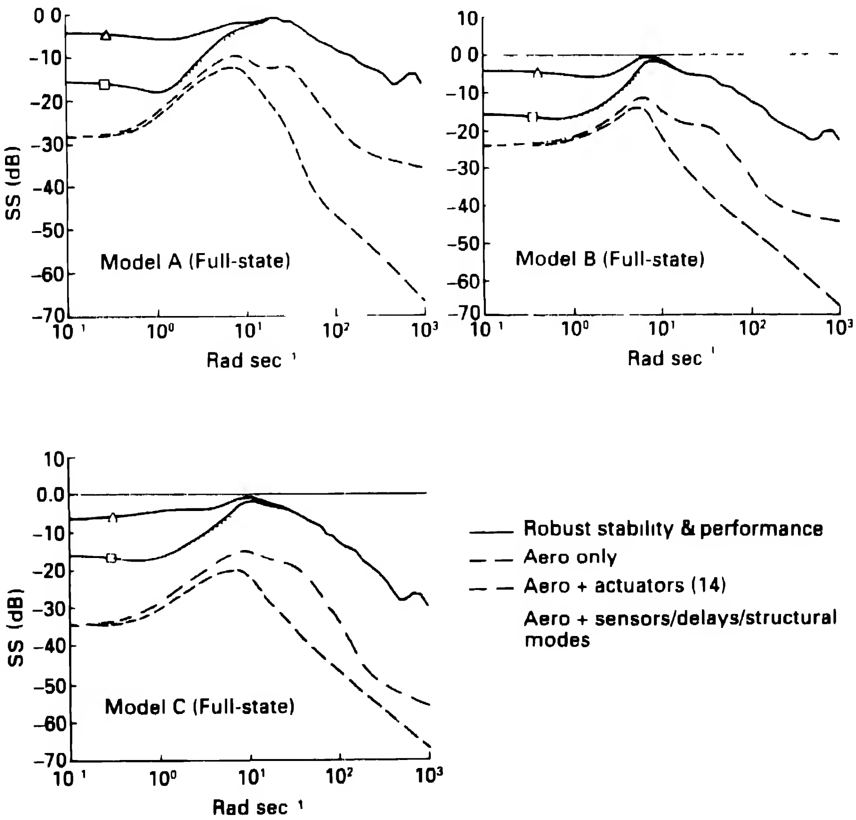


FIG. 12. SSV of total robust stability (□), individual robust stability (dashed), and total robust performance (△) against uncertainty types 1–3 (aerodynamic, actuators and sensors).

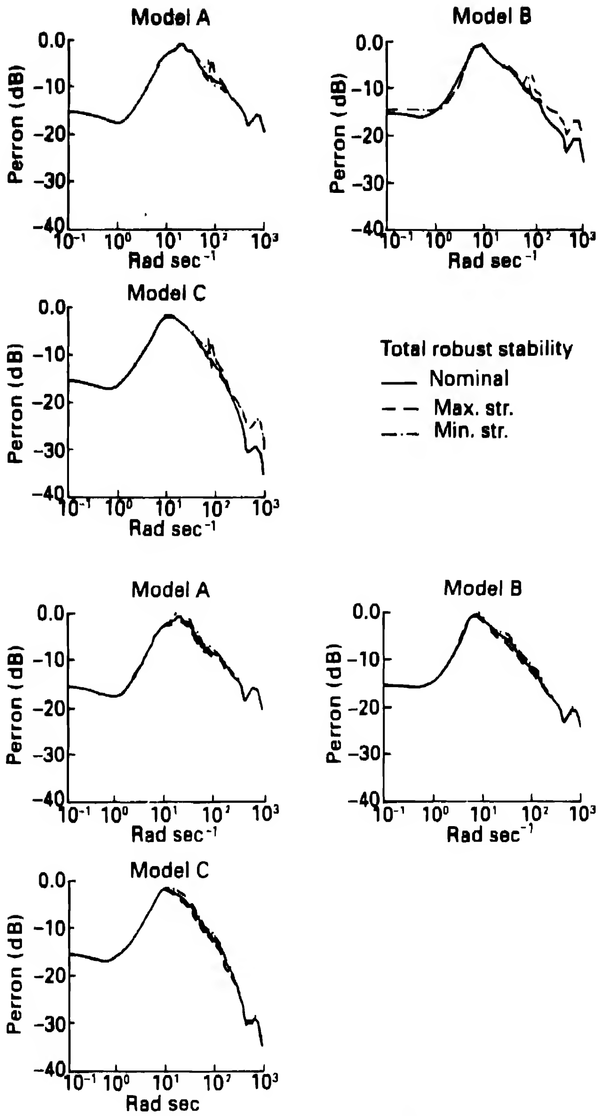


FIG. 13. SSV of total robust stability against uncertainty types 1–3 as aircraft weight varies.

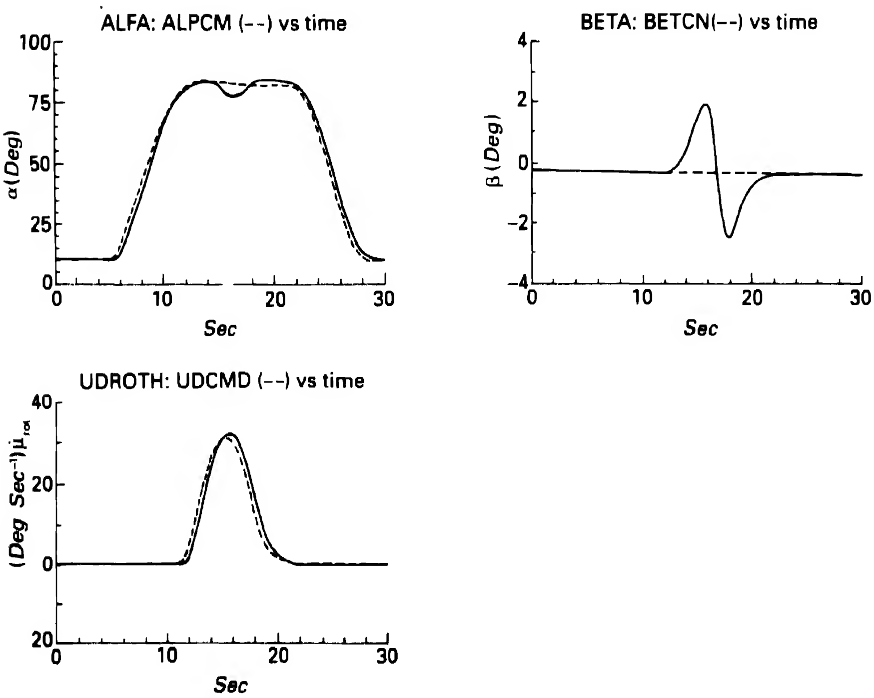


FIG. 14. Full model nonlinear simulation—angle-of-attack (α), side-slip (β), roll angle ($\dot{\mu}_{rot}$).

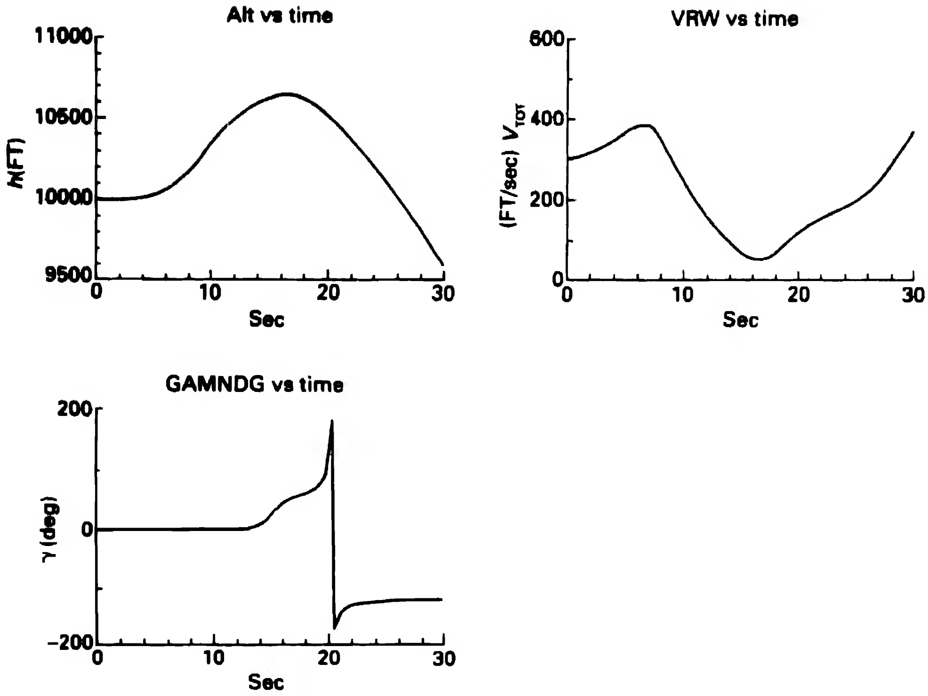


FIG. 15. Full model nonlinear simulation—altitude (h), total velocity (V_{tot}), heading angle (γ)

facts are observed from the nonlinear time history:

- The aircraft is stabilized by the fixed eight-state 3×3 H^∞ controller and the fixed LQR pure gain inner-loop controller throughout the maneuver without gain scheduling.
- The aircraft tracks the angle of attack (α) and roll commands ($\dot{\mu}_{rot}$) closely.
- Only two degrees peak sideslip angle (β) occur after roll command ($\dot{\mu}_{rot}$) is initiated.

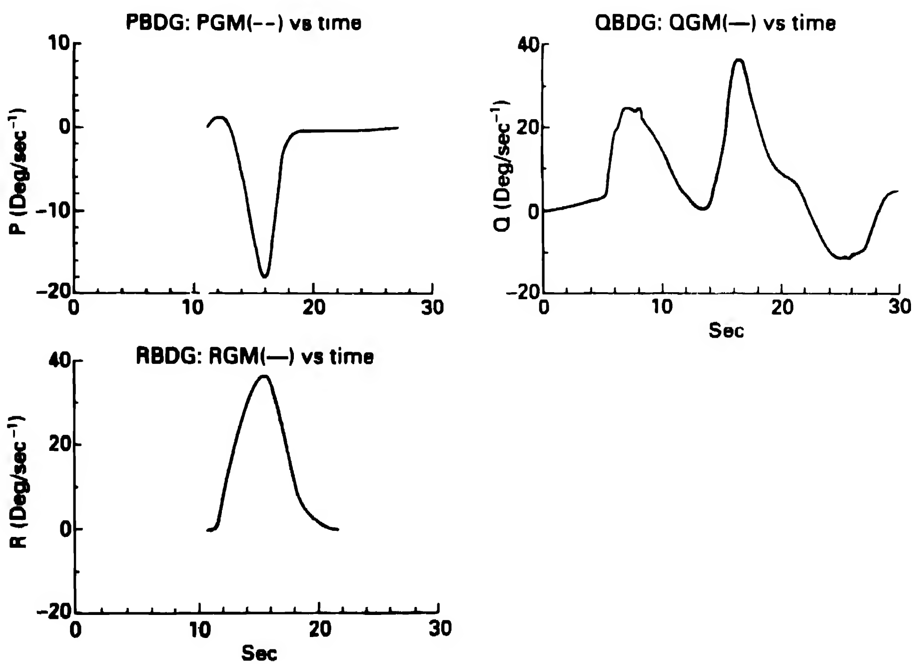


FIG. 16. Full model nonlinear simulation—angular rates (P, Q, R).

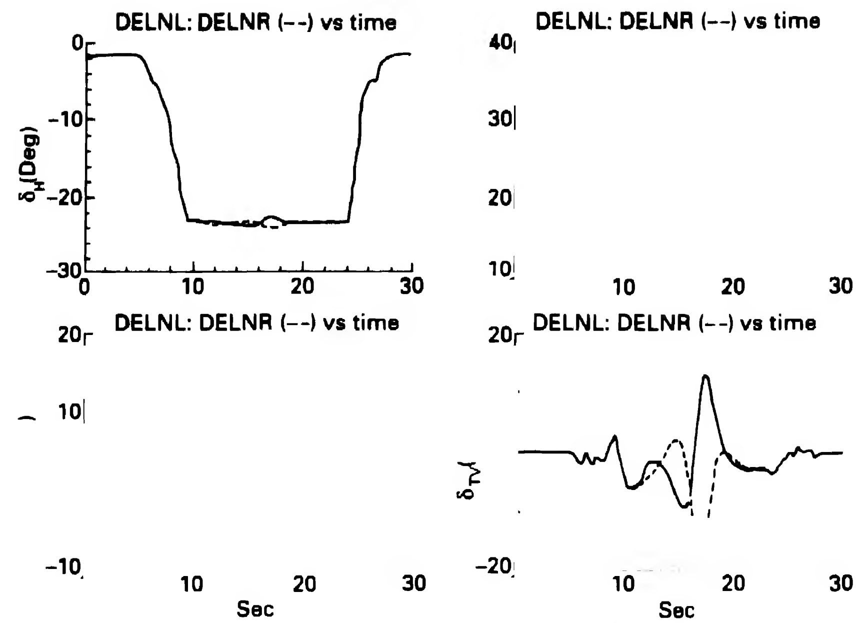


FIG. 17. Full model nonlinear simulation—horizontal tails (δ_H), leading edge flaps (δ_N), trailing edge (δ_F), thrust vectoring (δ_{TV}).

4. CONCLUSIONS

This paper presents a new approach involving the use of H^∞ theory for the design of a robust controller for a supermaneuverable fighter performing a Herbst-type maneuver. The fighter model is a modified F/A-18 Hornet with a pair of gimballed thrust vectoring nozzles. With the supermaneuverability provided by the thrust vectoring, the challenging Herbst-type maneuver, was selected. The uncertainty model and flying qualities specifications for this maneuver were identified and cast into the design process.

Both linear robustness analysis and full dynamics nonlinear simulation have shown the combined power of H^∞ and the robust equalization inner-loop. The use of the inner loop as an equalization loop for multi-model mixed sensitivity H^∞ control synthesis is an essentially new idea. This equalization property was the key to being able to design a fixed (i.e. not gain-scheduled) controller for which robust performance can be guaranteed throughout the wide range of plant time-variations associated with the Herbst maneuver. The role of inner-loop feedback in equalizing plants subject to large variations is in addition to, and should not be confused with, its important and better known role (Safonov *et al.*, 1990) in damping plant poles so as to avoid the robustness problems that result from the inevitable cancellation stable plant poles by zeros of H^∞ mixed-sensitivity controllers (see, for example, Sefton and Glover, 1990).

By combining mixed-sensitivity H^∞ control synthesis with the robust equalization inner-loop, we have been able to achieve a robust flight control design that not only can stabilize the aircraft throughout the entire maneuver in the face of anticipated structured and unstructured uncertainties and time-variations, but also possesses approximately decoupled tracking performance throughout the maneuver envelope. Most importantly, the new multi-model design concept presented here allowed the eight-state 3×3 H^∞ controller to be implemented without gain scheduling.

Acknowledgements—This project required the cooperation and support of a number of individuals from several disciplines. Their collective efforts are duly appreciated by the authors. Special thanks go to P. Shaw, D. Joshi, J. Kalviste and E. Tich for sharing their high α flight dynamics expertise and reviewing the first draft of the paper. We also thank Lt. Richard Adams, Capt. Sharon A. Heise, Capt. Mary Manning, Dr Hsi-Han Yeh, Dr Siva Banda, and John Bowlus, the Air Force contract monitors for their helpful technical comments and encouragement throughout the activity.

REFERENCES

Chiang, R. Y. and M. G. Safonov (1988). *Robust-Control Toolbox*. Math Works, Natick, MA.

Chiang, R. Y., M. G. Safonov and J. A. Tekawy (1990). H^∞ flight control design with large parameteric robustness . *Proc. of American Control Conf.*, San Diego, CA. IEEE Press, New York, pp. 2496–2501.

Doyle, J. C., K. Lenz and A. Packard (1987). Design examples using μ-synthesis: Space shuttle lateral axis FCS during reentry. NATO ASI Series, F34, *Modeling, Robustness and Sensitivity Reduction in Control Systems*, pp. 127–154. Springer-Verlag, Berlin.

Doyle, J. C., J. E. Wall and G. Stein (1982). Performance and robustness analysis for structured uncertainty. *Proc. 21st IEEE Conf. Decision Contr.*, Orlando, FL. IEEE Press, New York, pp. 629–636.

Falb, P. L., and W. A. Wolovich (1967). Decoupling in the

design and synthesis of multivariable control systems. *IEEE Trans. Aut. Control*, **AC-12**, 651–659.

Georgiou, T. T. and M. C. Smith (1990). Optimal robustness in the gap metric. *IEEE Trans. Aut. Control*, **AC-35**, 673–686.

Haiges, K., E. Tich and K. Madden (1989). Robust control law development for modern aerospace vehicles, Task 1: model development. Technical report No. WRDC-TR-89-3080, Wright–Patterson AFB, Dayton, OH.

Haiges, K., K Madden, B. Eller, A. Emami-Naeini, M. R. Anderson and T. L. Trankle (1990). Robust control law development for modern aerospace vehicles, Task 2: Control System criteria and specifications. Technical report No WRDC-TR-90-3005, Wright–Patterson AFB, Dayton, OH.

Haiges, K., R. Y. Chiang, K. Madden, A. Emami-Naeini, M. R. Anderson and M. G. Safonov (1991). Robust control law development for modern areospace vehicles, Final Report. Technical report, WRDC-TR-9-3105, Wright–Patterson AFB, Dayton, OH.

Herbst, W. A. (1980). Future fighter technologies. *J of Aircraft*, **17**, 561–566.

Hodgkinson, J. (1982). Equivalent systems criteria for handling qualities of military aircraft. AGARD-CP-333, Criteria for Handling Qualities of Military Aircraft, Fort Worth, TX.

Isidori, A., A. J. Krener, C. Gori-Giorgi and S. Monaco (1981) Nonlinear decoupling via feedback: A differential geometric approach. *IEEE Trans. Aut. Control*, **AC-26**, 331–345.

Kalviste, J. (1986). Spherical mapping and analysis of aircraft angles for maneuvering flight *AIAA Atmospheric Flight Mechanics Conf.*, Williamsburg, VA AIAA-86-2283, Washington DC

McRuer, D. T., I Ashkenas and Graham (1973) *Aircraft Dynamics and Automatic Control.* Princeton University Press, Princeton, NJ

MIL-STD-1797 (1987) *Flying Qualities of Piloted Vehicle*, Dept. of Defense, U.S.A. ASD/ENESS Wright–Patterson, AFB, Dayton, OH.

Pernebo, L. (1981). An algebraic theory for the design of controllers for linear multivariable systems—Parts I and II. *IEEE Trans. Aut. Control*, **AC-26**, 171–193.

Safonov, M. G. and M. Athans (1977). Gain and phase margin for multiloop LQG regulator. *IEEE Trans. Aut. Control*, **AC-22**, 173–178.

Safonov, M. G. and R. Y Chiang (1988a). CACSD using the state space L^∞ theory—a design example. *IEEE Trans. Aut. Control*, **AC-33**, 477–479.

Safonov, M. G. and R. Y. Chiang (1988b). Model reduction for robust control: A Schur multiplicative-error method. *Int. J of Adaptive Control and Signal Processing*, **2**, 259–272

Safonov, M. G., D. J. N. Limebeer and R. Y. Chiang (1989). Simplifying the H^∞ theory via loop shifting matrix pencil and descriptor concepts. *Int. J. Control*, **50**, 2467–2488.

Safonov, M. G , R. Y. Chiang and H. Flashner (1990). H^∞ robust control synthesis for a large space structure. *J. Guidance, Contr., and Dynamics*, **14**, 513–520.

Sefton, J. and K. Glover (1990). Pole/zero cancellations in the general H^∞ problem with reference to a two block design. *Systems and Control Letters*, **14**, 295–306.

Shaw, P. D., K. R. Haiges, S. R. Rock, J H. Vincent, A. Emami-Naeini, R. Anex, W S Fisk and D. F. Berg (1988). Design methods for integrated control systems. Report AFWAL-TR-88-2061, Wright-Patterson, AFB, Dayton, OH

Tekawy, J. A., M. G. Safonov and R. Y. Chiang (1989). Algorithms for computing the multivariable stability margin *Third Annual Conf. on Aerospace Computational Control*, Oxnard, CA, Vol. 1, pp 104–118.

Zames, G. (1981). Feedback and optimal sensitivity: Model reference transformations, multiplicative semi-norms, and approximate inverses. *IEEE Trans. Aut. Control*, **AC-26**, 301–320.

Automatica, Vol. 29, No. 1, pp. 129–145, 1993
Printed in Great Britain.

0005–1098/93 $6.00 + 0.00
Pergamon Press Ltd

Robust Integrated Flight/Propulsion Control Design for a STOVL Aircraft Using H-infinity Control Design Techniques*†

SANJAY GARG‡

Proper formulation of the H_∞ *optimal control synthesis problem is shown to lead to centralized controllers for integrated flight/propulsion systems that provide robustness to plant parameter variations and modelling uncertainties.*

Key Words—Robust control; robustness analysis; control system synthesis; centralized control; flight control; propulsion control; integrated plant control; control applications

Abstract—Results are presented from an application of H_∞ control design methodology to a centralized integrated flight/propulsion control (IFPC) system design for a supersonic Short Take-Off and Vertical Landing (STOVL) fighter aircraft in transition flight. The emphasis is on formulating the H_∞ optimal control synthesis problem such that the critical requirements for the flight and propulsion systems are adequately reflected within the linear, centralized control problem formulation and the resulting controller provides robustness to modelling uncertainties and model parameter variations with flight condition. Experience gained from a preliminary H_∞ based IFPC design study performed earlier is used as the basis to formulate the robust H_∞ control design problem and improve upon the previous design. Detailed evaluation results are presented for a reduced order controller obtained from the improved H_∞ control design showing that the control design meets the specified nominal performance objective as well as provides stability robustness for variations in plant system dynamics with changes in aircraft trim speed within the transition flight envelope.

INTRODUCTION

THE TREND IN future military fighter/tactical aircraft design is towards aircraft with new/enhanced maneuver capabilities such as Short Take-Off and Vertical Landing (STOVL) and high angle of attack performance. An integrated flight/propulsion control (IFPC) system is required in order to obtain these enhanced capabilities with reasonable pilot workload. An integrated approach to control design is then necessary to achieve an effective IFPC system. Such a design approach is currently being developed at NASA Lewis Research Center under an in-house research effort. This methodology is referred to as IMPAC—Integrated Methodology for Propulsion and Airframe Control (Garg *et al.*, 1991). The significant features of the IMPAC methodology are that it consists of first designing a centralized controller considering the airframe and propulsion systems as one integrated system and then partitioning the centralized controller into decentralized subsystem controllers for state-of-the-art IFPC implementation.

The major issue related to the centralized controller design portion of IMPAC is the choice of the control synthesis technique that "best" suits the IFPC objectives. Not only should the synthesis technique provide for means of formulating the design criteria for the centralized control design such that it adequately reflects the performance specifications of the "total" system, i.e. the airframe integrated with the propulsion system, but is should also result in controllers of reasonable complexity with guaranteed performance and robustness characteristics. Recent advances in H_∞ control theory (Doyle *et al.*, 1989; Safonov *et al.*, 1989) and computational algorithms to solve for H_∞ optimal control laws (Anon., 1989; Chiang and Safonov, 1988) have made this theory a viable candidate to be applied to complex multivariable control design problems. A preliminary investigation of the applicability of H_∞ control theory to the centralized feedback controller design portion of the IMPAC approach was conducted earlier via an example IFPC design study (Garg *et al.*,

* Received 1 October 1991; received in final form 2 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor P. Dorato.

† The original version of this paper was presented at the 1991 American Control Conference which was held in Boston, MA, U.S.A.

‡ Advanced Control Technology Branch, NASA Lewis Research Center, Cleveland, OH, U.S.A.

1990). The results reported in Garg *et al.* (1990) are quite encouraging in that they demonstrate that H_∞ control theory has the promise to provide the framework to meet the requirements of a centralized IFPC design. However, the control design of Garg *et al.* (1990) was a preliminary design and detailed evaluation of that design identified various areas in which the controller performance needed to be improved. The objectives of this paper are to develop insight into formulating a robust control problem with the H_∞ control design framework and to improve upon the control design of Garg *et al.* (1990). Towards this goal results are presented from an H_∞ based IFPC redesign for the linear model of a STOVL aircraft in transition flight considered earlier in Garg *et al.* (1990).

The paper is organized as follows. The vehicle models to be used for control design and evaluation are first discussed. The H_∞ control design is then presented along with some discussion of the formulation of the IFPC design objectives within the framework of the H_∞ control problem. The emphasis is on formulating the problem such that the resulting controller is robust to modelling uncertainties and parameter variations with changes in flight condition. Evaluation results are presented for a reduced order approximation of the improved H_∞ controller and improvements over the controller design of Garg *et al.* (1990) are demonstrated.

VEHICLE MODEL

The vehicle considered in this study is representative of the delta winged E-7D supersonic STOVL airframe powered by an enhanced version of a high performance military turbofan engine (Akhter *et al.*, 1989). The aircraft is equipped with the following controls: ejectors to provide propulsive lift at low speeds and hover; a 2D-CD (Two-Dimensional, Convergent–Divergent) vectoring aft nozzle with afterburner for supersonic flight; a vectoring ventral nozzle for pitch control and lift augmentation during transition; and jet reaction control systems (RCS) for pitch, roll and yaw control during transition and hover. A schematic diagram of the aircraft with relative location of the various control effectors mentioned above is shown in Fig. 1. Engine compressor bleed flow is used for the RCS thrusters and the mixed engine flow is used as the primary ejector flow. Detailed ducting diagrams of the engine and discussion of the ejector STOVL concept are available in Akhter *et al.* (1989).

Two separate computer simulations, one for the aircraft (six-degree-of-freedom aerodynamic

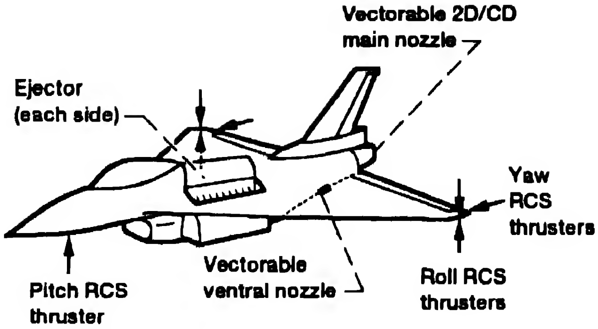


FIG. 1. Control effectors for E-7D aircraft.

model with steady-state engine performance model) and one for the propulsion system (aero-thermo dynamic model) were used to assess performance capabilities of the aircraft and to generate open-loop linear models for control design (Akhter *et al.*, 1989). The procedure for generating integrated airframe and propulsion models for control design and evaluation from the two separate simulations is discussed in Garg *et al.* (1990). The integrated linear design model used in this study is of the form

$$\dot{\mathbf{x}} = \mathbf{AX} + \mathbf{Bu}; \quad \mathbf{y} = \mathbf{Cx} + \mathbf{Du}, \tag{1}$$

where the state vector is

$$\mathbf{x} = [u, v, w, p, q, r, \phi, \theta, \mathrm{N2}, \mathrm{N25}, \mathrm{T41}, \mathrm{T3}, \mathrm{P6}]^T, \tag{2}$$

with

u = Axial velocity (ft s^{-1}),
v = Lateral velocity (ft s^{-1}),
w = Vertical velocity (ft s^{-1}),
p = Roll rate (rad s^{-1}),
q = Pitch rate (rad s^{-1}),
r = Yaw rate (rad s^{-1}),
ϕ = Roll attitude (rad),
θ = Pitch attitude (rad),
N2 = Engine fan speed (rpm),
N25 = High pressure compressor speed (rpm),
T41 = High pressure turbine inlet temp. (°R),
T3 = Combustor inlet temp. (°R),
P6 = Tailpipe entrance total pressure (psi),

the control inputs are

$$\mathbf{u} = [\delta e, \delta a, \delta r, \mathrm{AQR}, \mathrm{AYR}, \mathrm{ARR}, \mathrm{WF}, \mathrm{A8}, \mathrm{ETA}, \mathrm{A78}, \mathrm{ANG79}, \mathrm{ANG8}], \tag{3}$$

with

δe = Elevator deflection (deg)
δa = Aileron deflection (deg)
δr = Rudder deflection (deg)
AQR = Pitch RCS area (in^2)
AYR = Yaw RCS area (in^2)

ARR = Roll RCS area (in^2)
WF = Fuel flow rate (lbm hr^{-1})
A8 = Aft nozzle throat area (in^2)
ETA = Ejector butterfly angle (deg)
A78 = Ventral nozzle area (in^2)
ANG79 = Ventral nozzle vectoring angle (deg)
ANG8 = Aft nozzle vectoring angle (deg)

and the outputs are

$$\mathbf{y} = [V, \dot{V}, \theta, q, \gamma, \phi, p, \beta, r, \mathrm{N2}, \dot{\beta}], \quad (4)$$

with

V = True airspeed (ft s^{-1})
$\dot{V}$ = Acceleration along flight path (ft s^{-2})
γ = Longitudinal flight path angle (deg)
β = Sideslip angle (deg)
$\dot{\beta}$ = Rate of change of sideslip angle (deg s^{-1}).

The other outputs are as discussed under state description except that the angular positions and rates are in degrees. The integrated plant systems matrices, **A**, **B**, **C**, and **D**, are listed in the Appendix.

Some discussion of the choice of control inputs **u** for the linear design model is relevant here. The choice of outputs **y** is discussed later in the paper. The E-7D aircraft is equipped with left and right elevons on the trailing edge of the delta wing. Collective deflection of the elevons provides the classical elevator pitch control while differential use of the elevons provides the aileron roll control. So the elevator (δe) and aileron (δa) along with the rudder (δr) are used as the airframe control inputs in the design model. Only three RCS areas, AQR, AYR and ARR, are used as RCS control inputs in the linear design model whereas the full nonlinear-model has five controlled RCS areas. The reasons for this are that the nose pitch RCS and the two yaw RCS thrusters provide thrust in only one direction as shown in Fig. 1, and the wing tip roll RCS thrusters are to be used differentially for roll control and collectively for pitch control. For instance yaw RCS thrusters provide only forward thrust, so left yaw RCS is used for right yaw and right yaw RCS is used for left yaw in the nonlinear model. Using both left and right yaw RCS areas as control inputs in the design model can result in a control design that uses the two areas differentially to enhance yaw control which is inconsistent with the actual implementation. Similar reasons apply for pitch and roll RCS area selections. An RCS distribution logic that will distribute the three design model RCS commanded areas to the five actual areas in the nonlinear control implementation is shown in Fig. 2. Since the nose pitch RCS thruster only provides positive (pitch up)

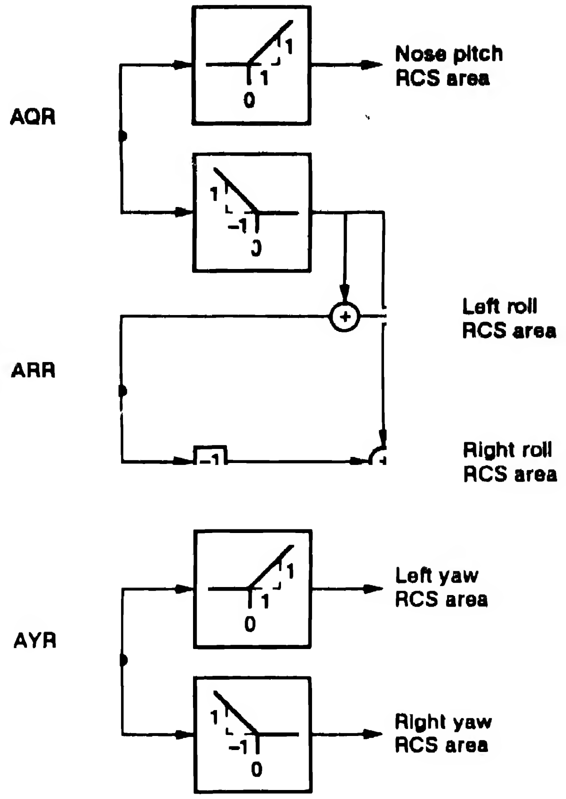


FIG. 2 RCS area distribution logic for control implementation.

pitching moment, a negative AQR command is distributed among the left and right wing tip RCS thrusters to generate the commanded negative pitching moment taking into account the relative pitch control effectiveness of the nose and wing tip RCS thrusters.

Figure 2 is presented here to clarify the discussion on choice of RCS areas for linear design model although the logic of Fig. 2 is not used to evaluate the control designs presented in this paper. Note that there is an absolute nonlinearity in the relationship from commanded RCS areas to compressor bleed flow demand in that although the RCS area may be positive or negative depending on the desired direction of RCS thrust, the compressor bleed flow demand (WB3) to generate the thrust is always positive. For a linear model trimmed about zero RCS areas, this relationship is of the form $WB3i = K_i\,|AiR|$ where $WB3i$ is the demanded bleed flow due to AiR command with i representing Q, Y or R for pitch, yaw or roll RCS, respectively, K_i is an appropriate constant and $|\cdot|$ represents absolute value. This nonlinearity was not taken into account in the linear control design of Garg *et al.* (1990) and the resulting controller was shown to lead to unacceptable performance when evaluated with this nonlinearity. In this paper, the critical absolute nonlinearity from RCS area commands to WB3 demand is

accounted for in the linear control design process by a proper formulation of the control design problem within the H_∞ framework. The details of this formulation are discussed later in the paper.

The control inputs, WF, A8, ETA, A78, ANG79 and ANG8 in the vector **u**, are the propulsion system inputs. The ejector butterfly valve angle (ETA) controls the engine airflow to the ejectors, thus providing a means of controlling ejector thrust. There are separate control valves for the left and right ejectors, however the two valve angles are set to be equal in the aircraft simulation because no test data are available on the differential use of the ejectors for roll control. Therefore only one butterfly valve angle is used as the control input in the design model. The other five propulsion system control inputs in the design model were just as defined in the full cycle-deck engine simulation (Akhter *et al.*, 1989).

The flight phase considered in this study is the decelerating transition during approach to hover landing. This phase presents a challenging control design problem becuase the control of the aircraft is transitioning from aerodynamically generated forces and moments to those generated by the propulsion system. For this study, three linear integrated models were obtained corresponding to steady-state level flight at trim speeds of $V_0 = 60$, 80 and 100 knots with trim flight path angle $\gamma_0 = -3$ deg for all three models. The trim strategy used to generate the linear models is as follows. (i) The aircraft pitch attitude was set at $\theta_0 = 7$ deg to provide adequate visibility during landing, and the elevon settings were chosen to correspond to elevator deflection $\delta e_0 = 20$ deg so that adequate elevator control authority is available for maneuvers during transition (ii) the aft nozzle vectoring angle was set to $ANG8_0 = 0$ so that the total vectoring authority is available for pitch control, and the ventral nozzle vectoring angle was set to $ANG79_0 = 64$ deg to ensure that the aircraft can be trimmed at the low speed (60 knots) with small but nonzero aft nozzle thrust, hence nonzero throat area. This trim strategy leaves adequate A8 and ANG79 control authority available for these controls to be used for active speed and pitch control in transition; (iii) the aircraft is then trimmed by using the three thrusts—aft nozzle, ventral nozzle and the combined ejectors, as the trim controls. The resulting trim values of A8, ETA, and A78 for the three trim speeds are listed in Table 1. Also listed in Table 1 are the trim values of fuel flow (WF) and engine fan speed (N2) which define the engine operating point. Note that the quantities shown in Table 1 are a percentage of the corresponding trim values for the 80 knot model. As seen from Table 1, thrust is transferred from aft nozzle to ventral nozzle as the aircraft slows down and the engine operating point is moved up to a higher gross thrust level (increased WF and N2) to generate the propulsive lift needed to compensate for the loss in aerodynamic lift.

Open loop frequency-domain and time-domain analyses of the three linear models indicated that the 80 knot model provides a "good average" of the three models in terms of input/output response behavior. Thus the 80 knot model was used for control design with the 60 and 100 knot models being used to evaluate the stability and performance robustness of the

TABLE 1. TRIM CONTROL FOR LINEAR MODELS (LISTED AS PERCENT (%) OF 80 KNOT MODEL TRIM VALUES)

V_0 (knots)	60	80	100
$A8_0$ (%)	11	100	268
ETA_0 (%)	96	100	118
$A78_0$ (%)	133	100	32
WF_0 (%)	116	100	71
$N2_0$ (%)	104	100	90

TABLE 2. EIGENVALUES OF LINEAR MODELS

Eigenvalue			Description
60	80	100	V_0 (knots)
-7.9×10^{-2}	-8.5×10^{-2}	-7.9×10^{-2}	Spiral mode
$-0.09 \pm j0.24$	$-0.11 \pm j0.28$	$-0.12 \pm j0.31$	Phugoid mode
1.05	1.29	1.44	Unstable SP†
−1.64	−1.73	−1.77	Roll mode
−1.74	−2.09	−2.45	Stable SP
$0.05 \pm j1.99$	$-0.23 \pm j2.27$	$-0.48 \pm j2.53$	Dutch roll
$-5.58 \pm j0.74$	−4.12	−3.33	Rotor speeds
	−7.11	−3.83	
−29.73	−29.39	−27.73	Air temps
−38.67	−38.21	−39.95	
−172.0	−199.7	−277.1	Pressure mode

† Short period.

nominal design, develop controller scheduling and evaluate off-design performance. The eigenvalues of the linear models for the three flight conditions are listed in Table 2 and the airframe modes are identified in terms of their "classical" interpretation (McRuer *et al.*, 1973). As seen from Table 2, the aircraft is unstable in pitch with the short period (SP) mode becoming less unstable as the trim speed decreases. Also the dutch roll mode damping is very low and it decreases with trim speed with the mode going unstable at 60 knots. The engine temperature and pressure modes are very fast compared to the engine rotor dynamics and aircraft modes. Ideally these fast modes will not be considered in the linear design model in order to reduce the complications in the control synthesis procedure. However, the approach taken in this study is to include these modes in the design model and reduce the complexity of the controller through order-reduction after initial controller synthesis.

H_∞ CONTROL DESIGN

Design methodology

The IFPC design problem discussed earlier was formulated as a command tracking, disturbance rejection problem within the framework of the general mixed sensitivity H_∞ control problem (Safonov and Chiang, 1988). The detailed block diagram for the H_∞ formulation of the IFPC feedback control design is shown in Fig. 3. In Fig. 3, $\mathbf{z}$ are the controlled variables and $\mathbf{z}_c$ are the corresponding reference commands. The three transfer functions that are of interest for such a problem are the sensitivity function $S(s)$, the complementary sensitivity function $T(s)$, and the control transmission function $C(s)$. These represent the closed-loop transfers from the reference commands and disturbances to tracking errors, controlled variables and commanded control inputs, respectively. In order to influence both the low-frequency and high-frequency properties of the closed-loop system it is desirable to find a controller $\mathbf{K}(s)$ which minimizes a weighted norm of a combination of these three transfer functions, i.e.

$$\min_{\text{Stabilizing } K(s)} \|H(s)\|_\infty \quad \text{with} \quad H(s) = \begin{bmatrix} W_S(s)\cdot S(s) \\ W_T(s)\cdot T(s) \\ W_C(s)\cdot C(s) \end{bmatrix}$$

where $\|H(s)\|_\infty = \sup_\omega\,(\sigma_{\max}[H(j\omega)])$.

The weighting functions W_S, W_T, and $W_C(s)$ are the "knobs" used by the control designer to "tune" the controller $\mathbf{K}(s)$ such that the design objectives are met. For instance choosing W_S to be large at low-frequency ensures good command tracking performance and choosing W_T to be large at high frequencies ensures robustness to high-frequency unmodelled dynamics. W_C are chosen to ensure that control actuation bandwidths and control rate and deflections are practically achievable.

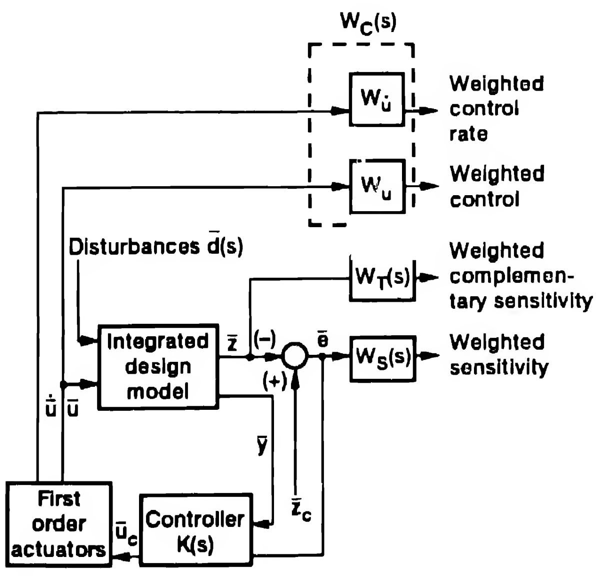


FIG. 3. Block diagram for H_∞ formulation of IFPC feedback design.

The H_∞ tracking formulation of Fig. 3 allows for feedback of plant measurements other than just tracking errors as inputs to the controller. This formulation then allows the simultaneous design of inner loop plant augmentation (stability or response shaping) and command tracking system. Such plant augmentation is an integral part of flight control design since the overall objective is to design a system for desired piloted handling qualities and not just an automatic command tracking system. Also, it has been shown in Chiang *et al.* (1990) that if a tracking problem is formulated within the H_∞ framework as purely a servo-mechanism problem, i.e. with controller inputs being just the tracking errors, then the H_∞ controller will be such that its transmission zeros cancel the stable poles of the design plant thus resulting in a closed-loop system that will have almost no robustness to plant parameter variations that result in change in open-loop pole locations. Allowing for feedback augmentation within the H_∞ control formulation overcomes this problem of lack of robustness as demonstrated in Garg *et al.* (1990).

Design specifications

The vectors $\mathbf{u}$ and $\mathbf{y}$ in Fig. 3 are the integrated design model inputs and outputs,

respectively, as discussed in the previous section. The controlled variables $\mathbf{z}$ were selected to be

$$\mathbf{z} = [Vv, Qv, \gamma, Pv, \beta, \mathrm{N2}]^T \tag{5}$$

with $Vv = \dot{V} + 0.1V$, $Qv = q + 0.3\theta$, $Pv = p + 0.1\phi$ and the others as discussed under plant outputs. This blending of controlled variables was chosen to provide the response types that are desirable for good handling qualities (Anon., 1970; Hoh and Mitchell, 1986) in transition flight. The choice of Vv corresponds to designing an acceleration (deceleration) command system with velocity hold, and the choice of Qv and Pv correspond to designing a rate command-attitude hold system. The break frequencies for switching from rate to attitude command for the case of Qv and Pv, and from acceleration to velocity command for the case of Vv, were chosen based on open-loop control effectiveness studies. For instance, the elevator (δe) is effective in pitch rate control in the frequency range of 0.3 rad s^{-1} to 10 rad s^{-1} and is effective in pitch attitude (θ) control for frequencies below 0.3 rad s^{-1}. The choice of γ in $\mathbf{z}$ provides for flight path angle control and the choice of N2 provides for tracking the fan speed commands generated by the engine operating schedule logic.

Designing the feedback controller $\mathbf{K}(s)$ to provide decoupled command tracking of the individual elements of $\mathbf{z}$ will result in a system that provides independent control of acceleration, pitch, flight path angle, roll and sideslip from the various pilot control effectors such as stick, throttle and rudder pedals, etc. thus reducing pilot workload, and also control of the propulsion system operating point (N2) independent of the aircraft motion. Independent control of roll (Pv) and sideslip angle (β) will result in a control system that provides automatic turn coordination thus further reducing pilot workload.

The choice of outputs $\mathbf{y}$, in equation (4), is consistent with the various quantities used for feedback in classical aircraft control augmentation (McRuer *et al.*, 1973). Rate feedback provides for improved damping while position (angles) feedback improves natural frequency of the appropriate mode.

Based on the experience gained from the control design study of Garg *et al.* (1990) an important criterion for control design is that the closed-loop system be robust to the absolute nonlinearity associated with the bleed flow demand from the RCS area commands—see earlier discussion. Also, the control design of Garg *et al.* (1990) was such that it did not further stabilize the dutch roll mode, i.e. the open-loop dutch roll mode was also a mode of the closed-loop system with the designed controller. This resulted in an unstable closed-loop system for the off-design integrated model at 60 knots due to the dutch roll mode being unstable at that flight condition. For this redesign effort, stabilization of the dutch roll mode to guarantee stable closed loop system over the transition flight conditions being considered was included as a design specification.

Control design

The design plant inputs and controlled outputs were normalized by maximum allowable deflections ($\mathbf{u}_{max}$) and maximum commanded values to be tracked ($\mathbf{z}_{cmax}$), respectively. The $\mathbf{u}_{max}$ were chosen to be reasonable deviations from the nominal (trim) values such that the total control deflection limits (as incorporated in the actuator models in the nonlinear simulation) will not be exceeded. Some of the $\mathbf{u}_{max}$ values corresponding to the propulsion system controls were further reduced to ensure that the total control usage will be within the limits imposed by safety requirements. For example, the allowable engine fuel flow for a given operating point is limited by the engine acceleration/deceleration schedule. The $\mathbf{z}_{cmax}$ were chosen based on handling qualities control requirements and open-loop control effectiveness analysis of the design plant to ensure that each element of $\mathbf{z}$ can be commanded individually to its maximum value within its frequency range of interest without exceeding $\mathbf{u}_{max}$ value for any of the control inputs. The $\mathbf{z}_{cmax}$ and $\mathbf{u}_{max}$ used in the control design are listed in the Appendix. The singular values of the scaled design plant for the six controlled variables defined in equation (5) are shown in Fig. 4. The fact that the minimum singular value in Fig. 4 is less then one implies that there exist combinations of numerical values of commands $\mathbf{z}_c$ such that although each element of $\mathbf{z}_c$ is less than its maximum value, the combined commands $\mathbf{z}_c$ cannot be tracked without exceeding the control limits $\mathbf{u}_{max}$ for some control input.

The sensitivity weights W_S and the complementary sensitivity weights W_T for each of the controlled variables were chosen to be first-order, to provide adequate frequency response shaping without overly increasing the resulting controller order. The W_S zero and pole for each controlled variable were chosen to result in a low-frequency gain of 1000, gain crossover frequency of 3–4 times the control bandwidth desired for good handling qualities for the controlled variable of interest, and a high-frequency gain of 0.1. Such a choice of W_S reflects the desire to synthesize a sensitivity

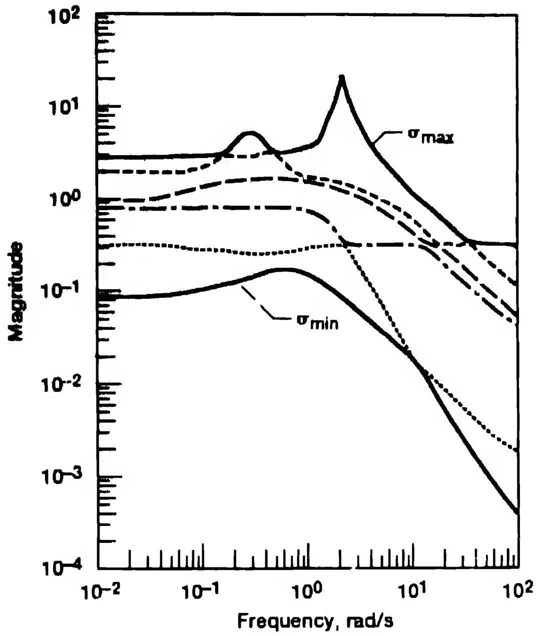


FIG 4 Singular values of scaled design plant

function which gives good steady-state tracking performance in the presence of disturbances and low-frequency modelling errors, good tracking performance up to the desired bandwidth of control and reduced emphasis on tracking at high frequencies where there are significant modelling errors and uncertainties. The W_T were chosen to obtain a low-frequency gain of 0, gain crossover frequency of about 1.2 times the corresponding W_S gain crossover frequency, and a high-frequency gain of 100. This choice of W_T ensures that the plant outputs are not penalized at low frequencies where command tracking is to be emphasized while at high frequencies the plant outputs are penalized heavily to provide controller gain attenuation for robustness to high-frequency unmodelled dynamics. The sensitivity and complimentary sensitivity weights for each controlled variable z_i are listed in the Appendix.

As part of the control weighting W_C, both the control inputs and control rates were weighted with W_u in Fig. 3 chosen to be the inverse of $\mathbf{u}_{max}$ and $W_{\dot{u}}$ chosen to be inverse of maximum control rate for each control input. The $\dot{\mathbf{u}}_{max}$ values used in the control design are listed in the Appendix. Since using full order actuator models for each control input would have resulted in a very high-order controller, first order actuator approximations were used in the control design. Describing function analysis (Ogata, 1970) of the full order actuators was first performed to determine the degradation in actuator bandwidth due to rate limiting when control commands corresponding to $\mathbf{u}_{max}$ are used at all frequencies. The worst-case rate-limited actuator bandwidth was then used as the bandwidth for first-order design actuators. For example, the pitch RCS area actuator bandwidth is 20 rad s^{-1} but the rate limit is 3.0 in s^{-1} which results in a worst case bandwidth of 6.0 rad s^{-1} for AQR_{max} of 0.7 in^2. Using such an approximation for the actuators in the design guarantees performance robustness and stability in the presence of actuator rate limiting. The actuator model numerical data are listed in the Appendix.

The bleed flow demand from the RCS acts as a disturbance on the engine operating point. In the propulsion system simulation and the linear model generated from this simulation, bleed flow appears as an external disturbance. However, when the airframe and propulsion models are integrated there is no explicit dependence on bleed flow since the bleed flow is generated from the RCS area commands which are the inputs to the integrated model. As pointed out earlier, the linear design model does not account for the absolute nonlinearity from the RCS area commands to bleed flow demand, so the effect of the RCS areas on the engine states as it appears in the integrated linear design model is erroneous. The linear design model was modified such that the elements of the control effectiveness matrix, **B**, from the RCS areas to engine states are zero, i.e. the RCS areas do not affect the engine operating point, and bleed flow was added as an external disturbance affecting the engine dynamics as in the linear engine model. For H_∞ control synthesis, the bleed flow disturbance was modelled as a scaled exogenous input with a scaling factor of 7 lbm s^{-1}, which is the maximum possible RCS bleed flow, and was considered to be a measurement available for feedback. Although bleed flow is not measurable, it can be estimated as a function of RCS areas and ambient and bleed duct pressure conditions or the bleed flow feedback can be implemented as nonlinear feedback of the RCS area commands. Since the RCS is used for control of aircraft angular rates, the bleed flow disturbance was filtered by a first order filter with a bandwidth of 3.5 rad s^{-1} which is representative of desired angular rate control bandwidths.

A procedure to account for plant model uncertainties within the framework of the H_∞ control problem is to use exogenous disturbance inputs which mimic the effects of model variations (Reichert, 1989). Such an approach was used in the current design study to provide enhanced lateral/directional stability by appropriate damping of the Dutch roll mode and robustness to changes in the lateral/directional dynamics with decreasing trim airspeed. Side

force, rolling moment and yawing moment changes were represented in the form of scaled exogenous disturbances in lateral acceleration ($\dot{v}$), roll acceleration ($\dot{p}$), and yaw acceleration ($\dot{r}$) with scaling factors of 0.5 ft s^{-2}, 0.1 rad s^{-2}, and 0.1 rad s^{-2}, respectively. The overall H_∞ control synthesis problem was then formulated as that of command tracking with bleed flow and lateral/directional disturbance rejection.

The design plant as discussed above is of 38th order consisting of the 13th order integrated airframe/propulsion system design model, first order actuators for the 12 control inputs, first order sensitivity and complementary sensitivity weights for the six controlled variables, and the first order filter for the bleed flow disturbance. So the H_∞ controller using the algorithm of Doyle *et al.* (1989) will be of 38th order. The H_∞ controller was obtained using a modified (by the author) version of the HINF_CONT function of the Robust Control Module software developed by Integrated Systems Inc. (see Anon., 1989). The H_∞ control design results with this design plant are shown in Fig. 5 in terms of the closed loop maximum singular values of the combined weighted functions $\sigma_{max}[H(j\omega)]$, weighted errors $\sigma_{max}[\mathbf{W}_S\mathbf{e}(j\omega)]$, weighted controlled outputs $\sigma_{max}[\mathbf{W}_T\mathbf{z}(j\omega)]$, weighted controls $\sigma_{max}[\mathbf{W}_u\mathbf{u}(j\omega)]$ and weighted control rates $\sigma_{max}[\mathbf{W}_{\dot{u}}\dot{\mathbf{u}}(j\omega)]$ with the commands $\mathbf{z}_c$ as inputs. $\|H(j\omega)\|_\infty = 10$ is achieved for this controller as seen from Fig. 5. In general a control design with $\|H(j\omega)\|_\infty \leq 1$ ensures that all the design specifications that are formulated through the various weightings will be met. However, this was not the case in the present design, because as stated earlier the minimum singular value of the scaled design plant itself was much less than one. The fact that control efforts greater than $\mathbf{u}_{max}$ will be required to track some combinations of commands is evident from the large (>1) maximum singular values of weighted controls at low frequencies. The fact that maximum singular value of weighted errors is greater than one at low frequencies indicates the performance/control trade-off made in the H_∞ minimization procedure. The fact that the maximum singular values of the weighted controlled outputs and control rates are less than one over all frequencies indicates that the H_∞ controller provides adequate "loop gain" roll-off for the closed-loop system to be robust to unmodelled high-frequency dynamics and that the control rate limits will not be exceeded for any combinations of commmanded variables. Further evaluation results are presented in the following section with reference to a reduced-order controller obtained from the H_∞ controller.

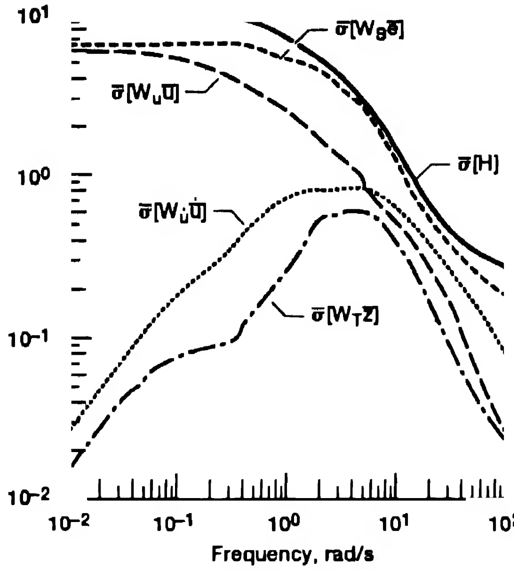


Fig. 5. Closed-loop weighted norms for H_∞ controller.

CONTROLLER REDUCTION AND EVALUATION

The controller order was reduced from 38 to 14 using a combination of modal residualization (Kokotovic and Sannvie, 1968) and balanced realization (Moore, 1981) reduction techniques. The order was first reduced from 38 to 22 by modal residualization of the high-frequency controller modes. The 22nd order controller was reduced to 16th order by a balanced realization approximation which was then reduced to 14th order again by residualization of the high-frequency modes. The state-space system matrices for the 14th order controller are listed in the Appendix. The eigenvalues of the final reduced order controller were bounded by $|\lambda| < 10$ thus indicating that digital implementation of the controller can be achieved with reasonable sampling rates. The performance with the full order and reduced order controllers is compared in Fig. 6 in terms of the maximum and minimum singular values of the closed-loop tracking system, $\sigma_{max}[\mathbf{T}(j\omega)]$ and $\sigma_{min}[\mathbf{T}(j\omega)]$ with $\mathbf{z}(s) = \mathbf{T}(s)\mathbf{z}_c(s)$. Clearly, there is no significant loss in tracking performance with the reduced order controller as seen by the excellent match for the closed-loop singular values up to a frequency of 20 rad s^{-1}. Note that other modern control reduction techniques such as frequency-weighted balanced realization (Enns, 1984) were not used because it was not clear how to apply these techniques for a controller structure which includes external disturbances as inputs (bleed flow in this particular case).

Extensive frequency-domain and time-domain analyses were performed to evaluate the closed-loop performance and robustness with the linear design model. All these analyses indicated that the reduced-order controller provides

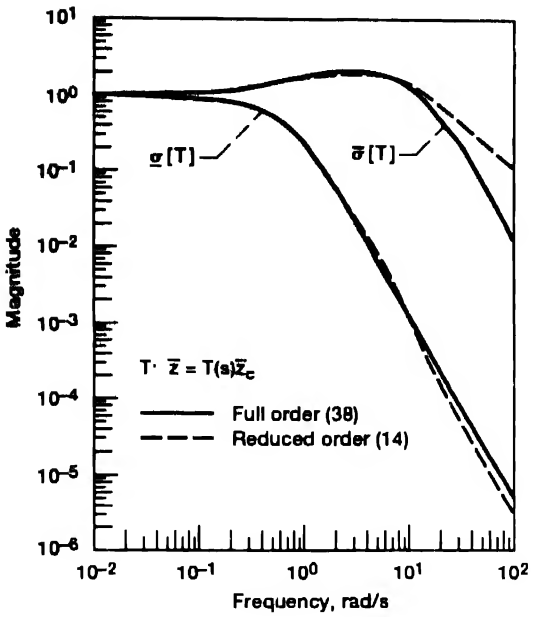


FIG 6. Closed-loop tracking singular values with full and reduced-order controllers

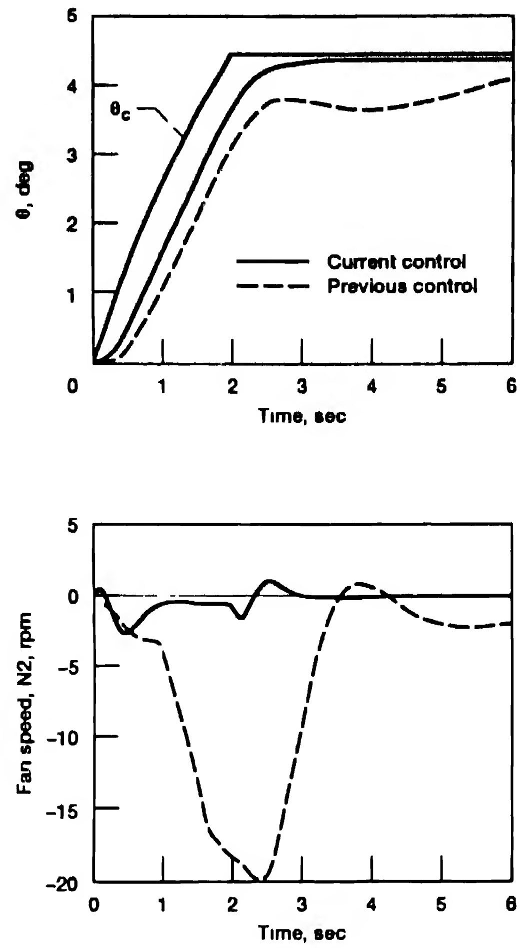


FIG 7 Pitch tracking performance in the presence of bleed flow nonlinearity

decoupled command tracking bandwidths and reasonable control actuation requirements. Note that although the open-loop design model is decoupled in the longitudinal and lateral/directional dynamics, coupling between these axes is introduced through the use of RCS control. For instance, using the yaw RCS will result in change in velocity because the yaw RCS generates axial force. So the fact that the control design achieves decoupled response in longitudinal and lateral/directional axes while using RCS for active control is significant.

The performance of the control design in the presence of the absolute nonlinearity from commanded RCS areas to compressor bleed flow demand was evaluated by including these nonlinear effects in the linear 80 knot design model. The numerical data for this absolute nonlinearity is listed in the Appendix. An example result is shown in Fig. 7 in terms of the closed-loop pitch attitude and engine fan speed response to a pitch variable (Qv) command. The Qv command was chosen to correspond to a transient pitch rate command and a steady-state pitch attitude command. Also shown in Fig. 7 is the corresponding closed-loop system response with the control design of Garg *et al.* (1990). The current design provides the desired improvement over the control design of Garg *et al.* (1990) in that it maintains good pitch tracking and fan speed regulation performance in the presence of the absolute bleed flow nonlinearity. Thus the formulation of the H_∞ control problem to reject bleed flow disturbance with the modified design plant, the modification being the removal of the RCS area effect on the engine dynamics, was successful in providing performance robustness to the bleed flow absolute nonlinearity.

In the transition flight phase, typical pilot control tasks are acceleration/deceleration to a desired airspeed while maintaining flight path angle or change to a desired flight path angle while maintaining airspeed. The IFPC design performance for these two typical tasks is shown in Figs 8 and 9. From Fig. 8(a), we note that the controller provides good command tracking of the transient acceleration and steady-state airspeed hold command while maintaining good regulation of pitch attitude, flight path angle and engine fan speed. The propulsion system actuation requirements, shown in Fig. 8(b), are quite reasonable and well within the limits established for the "small perturbation" linear control design (as seen in comparison with the values used for control and control rate normalization—see data in the Appendix). The resulting propulsion system generated thrusts are

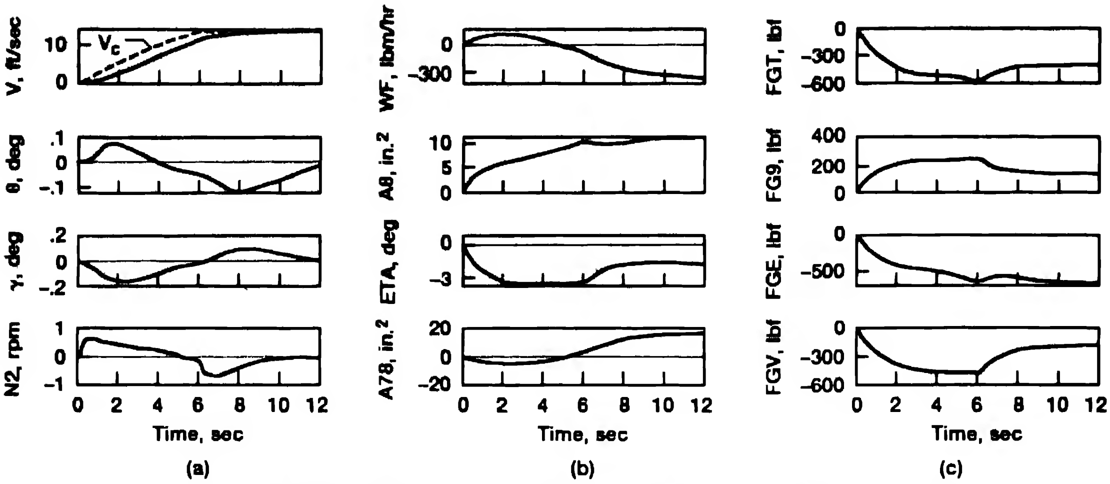


FIG. 8. Closed-loop response for transient acceleration/steady-state velocity hold command.

shown in Fig. 8(c) where

FGT = Total propulsion system gross thrust (lbf)
FG9 = Gross thrust from aft nozzle (lbf)
FGE = Gross thrust from ejectors (lbf)
FGV = Gross thrust from ventral nozzle (lbf).

FGT is an estimate for the total thrust that would be generated by the propulsion system if all the airflow were to exit by the aft nozzle at some specified nominal ambient conditions. As the aircraft speed increases, the aerodynamic lift will increase with a corresponding increase in drag. This will result in decreased requirement on propulsion system generated vertical thrust and an increase in the required axial thrust. The thrust responses of Fig. 8(c) are consistent with these requirements. The control design presented here corresponds to the "inner-loop" of an IFPC system. In actual implementation, there will be an "outer-loop" which will generate commands for the engine fan speed (N2) based on the total thrust requirements (FGT) to keep the engine operating on a desired schedule. From Fig. 8(c) we note that during the transition flight phase, an *acceleration* command for the aircraft results in decreased total propulsive thrust requirements and hence corresponds to an actual *deceleration* command for the engine rotor speeds.

Figure 9(a) shows that the control design provides good tracking of flight path angle command with decoupling from velocity, pitch attitude and fan speed. The propulsion system control actuation requirements are reasonable as seen from Fig. 9(b). As is to be expected, Fig. 9(c) shows an increase in ejector thrust to support the aircraft climb rate and a corresponding increase in total propulsion system thrust generation requirement.

Another important design criterion discussed earlier was to provide stability robustness to changes in the lateral/directional dynamics specially with reference to the low damping of

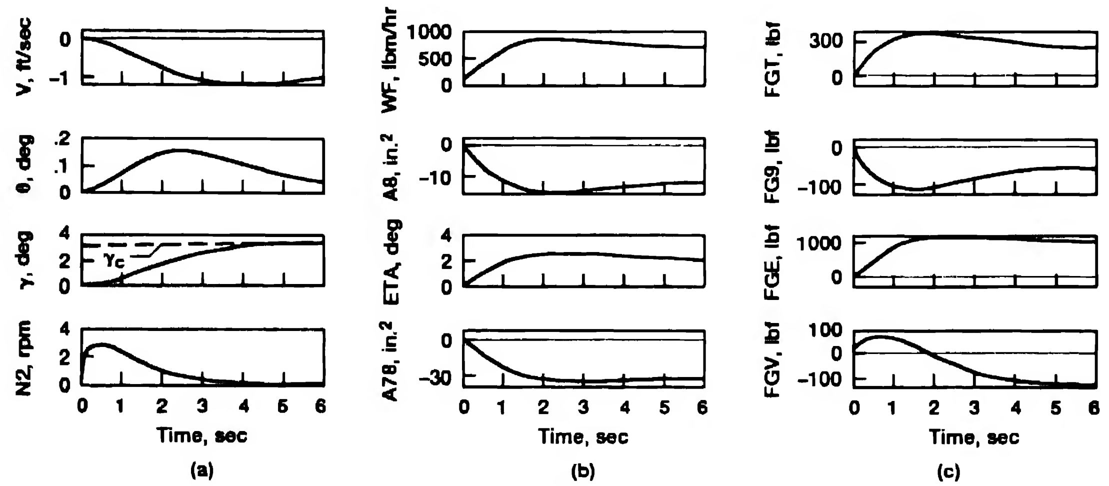


FIG. 9. Closed-loop response for step flight path angle command.

the dutch roll mode. For the current control design, the open-loop dutch roll mode was no longer a closed-loop mode with the 80 knot design plant. Moreover, all the closed-loop modes were well damped with the smallest damping ratio being 0.4. Detailed stability robustness evaluation of the control design to variations in the plant system **A** matrix due to changes in trim speed were conducted using structured singular value robustness tests (Doyle, 1985). The procedure for this stability robustness evaluation is discussed in the following.

First the variations in the airframe portion of the plant **A** matrix with changes in trim speed were identified. With the 80 knot model **A** matrix as the nominal, 13 elements of the airframe portion of the **A** matrix showed a change of 25% or greater in going from 80 to 60 knots or 80 to 100 knots. These elements correspond to change in axial (x-axis) force due to pitch rate, side force (y-axis) due to roll rate and yaw rate, vertical force (z-axis) due to axial speed (u), vertical speed (w) and pitch rate, rolling moment due to roll rate and yaw rate, pitching moment due to axial speed and pitch rate, and yawing moment due to side velocity (v), roll rate and yaw rate. Also the percentage change in the **A** matrix elements was approximately symmetrical about the nominal **A** matrix. For a given trim speed, V_0, these 13 elements of the **A** matrix were then modelled using the multiplicative uncertainty form as

$$A_{ij} = A_{ij}^{80}(1 + K_V^{ij}\Delta V), \qquad (6)$$

where A_{ij}^{80} is the i,j element of the 80 knot model **A** matrix, $\Delta V = V - 80$, with units in knots, and K_V^{ij} were chosen to be the gradient with respect to velocity for the $A(i,j)$ element from the 60 knot to 100 knot models, i.e. $K_V^{ij} = (A_{ij}^{100} - A_{ij}^{60})/40$. The numerical values for K_V^{ij} are listed in the Appendix. With such an uncertainty model, the perturbation $\Delta(s)$ for robustness analysis is of the form $\Delta(s) = \Delta V \cdot \mathbf{I}$ where $\mathbf{I}$ is an identity matrix of dimension 13, and an upper bound for structured singular value stability robustness measure considering ΔV to be *complex* is given by Apkarian (1989)

$$\mu = \sup\,[\rho(\mathbf{M}(j\omega))], \qquad (7)$$

where $\rho[\mathbf{A}]$ is the spectral radius of matrix **A** (maximum absolute eigenvalue of **A**) and $\mathbf{M}(s)$ is the interconnection matrix for robustness analysis which represents the nominal closed-loop system taking into account the structure of $\Delta(s)$. A detailed discussion of creating the interconnection matrix for this particular study is beyond the scope of this paper. The reader is referred to Doyle (1985) for a theoretical discussion of the subject of robustness analysis for structured uncertainties and to Anon. (1989) for a detailed exposition on creating the interconnection matrix for various forms of uncertainty models. The structured singular value for **A** matrix uncertainty as modelled by equation (6) is shown in Fig. 10. Also shown in Fig. 10 is the corresponding result with the control design of Garg *et al.* (1990). From Fig. 10, $\mu = 0.02$ for the current control design which implies that for the **A** matrix uncertainties as modelled, the closed-loop system will remain stable for velocity perturbations of up to $\Delta V = 1/\mu = 50$ knots. For the design of Garg *et al.* (1990), $\mu = 0.065$ which implies that the closed loop system stability with that control design would have been guaranteed only for velocity perturbations up to 15 knots. In fact, for $\Delta V = 15$ knots as modelled, the open-loop **A** matrix is such that the dutch roll mode is just unstable and the control design of Garg *et al.* (1990) was found to be unstable for this perturbation. The closed-loop system with the current control design was stable for the full off-design 60 and 100 knot plant models thus indicating that the use of exogenous disturbance inputs in the H_∞ problem formulation was successful in building-in robustness to variations in the lateral/directional dynamics.

Since the design plant is an integrated flight propulsion system, robustness of the design closed-loop system to changes in the propulsion

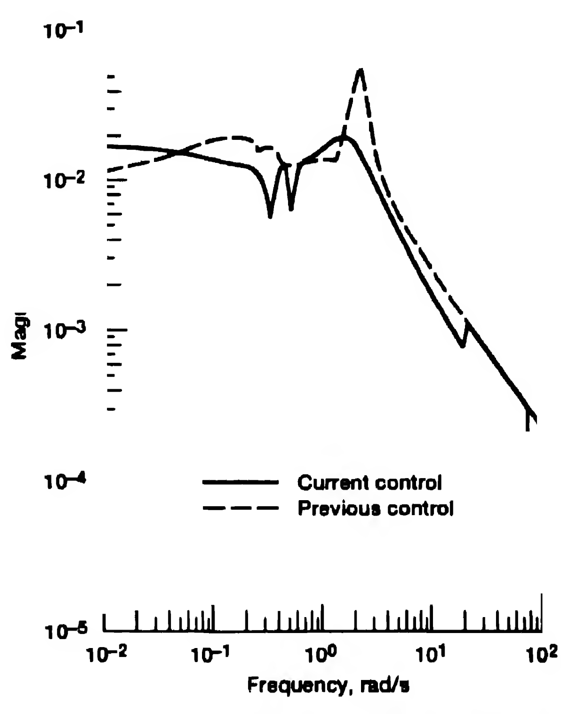


FIG. 10. Stability robustness measure $\rho(\mathbf{M})$ for variations in airframe portion of plant **A** matrix.

system dynamics is also an important criterion to be met. As mentioned in Sobey and Suggs (1963) the engine acceleration (fan/compressor speed changes) response to fuel flow characteristics can vary over a wide range due to the cumulative effect of various variables such as wear and differences in manufacturing tolerances, variable geometry setting, inlet distortion, environmental conditions, etc. For the purposes of robustness analysis, this variation in propulsion dynamics can be represented as uncertainty in the **A** matrix and **B** matrix elements corresponding to the change in rotor speed rates ($\dot{N}2, \dot{N}25$) with change in rotor speeds and fuel flow, respectively i.e. $A(i, j)$ and $B(i, k)$ with i and $j = 9$, 10 and $k = 7$ in the design model of equation (1). Structured singular value robustness with the 80 knot design model and the current controller design was performed considering independent multiplicative perturbations in these six elements of the **A** and **B** matrices. This analysis gave $\mu = 1.276$ which implies a stability margin $SM = 1/\mu = 0.79$, i.e. the closed-loop system will remain stable for changes of up to 79% from the 80 knot model in these **A** and **B** matrix elements. Thus the control design provides robustness to large variations in the engine dynamic response characteristics. The actual change in these elements between the design model and the off-design 60 and 100 knot models was much less than 79% thus indicating that the design controller will provide robustness to changes in engine dynamics due to changes in the engine operating point within the transition envelope.

Note that the robustness results presented so far are for *stability* robustness and not *performance* robustness. For the STOVL aircraft in transition flight, the control effectiveness of the various effectors changes significantly with speed. As speed decreases, the effectiveness of the aerodynamic control surfaces decreases and the control of aircraft attitude angles and angular rates is transferred from aerodynamic surfaces to propulsion system generated moments via RCS and nozzle vectoring controls. In general, it is very difficult to provide *performance* robustness to control effectiveness changes while maintaining nominal performance unless one is willing to accept very complex and high-order controllers. Controller scheduling is then required to provide satisfactory performance over the design flight envelope. A simple scheduling scheme of the form

$$\mathbf{K}(s) = \mathbf{K}_S \cdot \mathbf{K}^0(s),$$

where $\mathbf{K}(s)$ is the scheduled controller, $\mathbf{K}_S$ are scheduling gains which are a function of airspeed (V), and $\mathbf{K}^0(s)$ is the nominal controller (14th order for the 80 knot design point), was developed for the STOVL aircraft IFPC design (Garg and Ouzts, 1991). This scheduling scheme exploits the robustness characteristics of the nominal collector and was found to lead to satisfactory performance over the transition flight envelope.

The robust IFPC controller discussed in this paper is currently being partitioned into separate airframe and propulsion system controllers for ease of implementation and to allow for independent closed-loop propulsion system validation. The nonlinear safety and limit logic aspects of the propulsion system design are also being considered (Garg and Mattern, 1991). Piloted simulation evaluation of the overall controlled system, is planned in the near future.

CONCLUSIONS

The Integrated Flight/Propulsion Control (IFPC) system design presented in this paper demonstrates the applicability of an H_∞ control synthesis technique to integrated control design for complex systems such as Short Take-Off and Vertical Landing (STOVL) aircraft. The capability to address closed-loop performance, robustness and control actuation trade-offs within the framework of an H_∞ control problem formulation was demonstrated. In particular, the H_∞ control design for the IFPC problem, as formulated herein, provided robustness to the absolute nonlinearity associated with the compressor bleed flow demand from the Reaction Control System and to changes in the plant system dynamics over the transition flight envelope while achieving acceptable nominal performance. The IFPC design corresponds to a highly advanced pilot control mode for the STOVL aircraft wherein the pilot has independent direct control of flight path angle and acceleration along flight path. The integrated control makes the propulsion system thrust magnitude and vectoring management transparent to the pilot thus resulting in reduced pilot workload. Robust control design also reduces the requirements on controller scheduling thus simplifying control implementation.

Acknowledgement—The author would like to thank Dr Duane Mattern for his assistance in generating the linear models used in this study and for his constructive review comments on this manuscript.

REFERENCES

Akhter, M. M., J. H. Vincent, D. F. Berg and D. S. Bodden (1989). Simulation development for U.S./Canada controls technology program. *Proc. 20th Annual Modeling and Simulation Conference*, University of Pittsburgh, Pittsburgh, PA.

Anon. (1970). Military specification—flying qualities of piloted V/STOL aircraft. *MIL-F-83300*, Wright–Patterson AFB, OH.

Anon. (1989). *MATRIX$_X$ Robust Control Module—User's Guide*. Integrated Systems Inc., Santa Clara, CA.

Apkarian, P. R. (1989). Structured stability robustness improvement by eigenspace assignment techniques: A hybrid methodology. *J. of Guidance, Control and Dynamics*, **12**, 162–168.

Chiang, R. Y. and M. G. Safonov (1988). *Robust Control Toolbox—For Use With MATLAB—User's Guide*. The Mathworks, Inc., Natick, MA.

Chiang, R. Y., M. G. Safonov and J. A. Tekawy (1990). H_∞ flight control design with large parametric robustness. *Proc.* 1990 *ACC*, pp. 2496–2501.

Doyle, J. C. (1985). Structured uncertainty in control system design. *Proc.* 1985 *CDC*, pp. 260–265.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State-space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Aut. Control*, **AC-34,** 831–847.

Enns, D. F. (1984). Model reduction for control system design. Ph.D. dissertation, Dept. of Aeronautics and Astronautics, Stanford University, Stanford CA.

Garg, S. and D. L. Mattern (1991). Application of an integrated flight/propulsion control design methodology to a STOVL aircraft. 1991 *AIAA Guidance, Navigation and Control Conference*. Paper not published in proceedings. Available from AIAA as Paper No. 91-2792 or from NASA as NASA TM 105254.

Garg, S. and P. J. Ouzts (1991). Integrated flight/propulsion control design for a STOVL aircraft using H-infinity control design techniques. *Proc.* 1991 *ACC*, pp. 568–576.

Garg, S., D. L. Mattern, M. M. Bright and P. J. Ouzts (1990). H-infinity based integrated flight/propulsion control design for a STOVL aircraft in transition flight. *Proc.* 1989 *AIAA Guidance, Navigation and Control Conf.*, pp. 183–193.

Garg, S., P. J. Ouzts, C. F. Lorenzo and D. L. Mattern (1991). IMPAC—An integrated methodology for propulsion and airframe control. *Proc.* 1991 *ACC*, pp. 755–760.

Hoh, R. H. and D. G. Mitchell (1986). Proposed revisions to MIL-F-83300 V/STOL flying qualities specification. NADC-82146-60, Naval Air Development Center, Warminster, PA.

Kokotovic, P. V. and P. Sannuti (1968). Singular perturbation method for reducing model error in optimal control design. *IEEE Trans. Aut. Control*, **AC-13,** 377–384.

McRuer, D. T., I. Ashkenas and D. Graham (1973). *Aircraft Flight Dynamics and Automatic Control*. Princeton University Press, Princeton, NJ.

Moore, B. C. (1981). Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Aut. Control*, **AC-26,** 17–31.

Ogata, K. (1970). *Modern Control Engineering*. Prentice-Hall, NJ.

Reichert, R. (1989). Application of H-infinity control to missile autopilot design. *Proc.* 1989 *AIAA Guidance, Navigation and Control Conf.*, pp. 1065–1072.

Safonov, M. G. and R. Y. Chiang (1988). CACSD using the state space L_∞ theory—A design example. *IEEE Trans. Aut. Control*, **AC-33,** 477–479.

Safonov, M. G., D. J. N. Limebeer and R. Y. Chiang (1989). Simplifying the H_∞ theory via loop-shifting, matrix-pencil and descriptor concepts. *Int. J. of Control*, **50,** 2467–2488.

Sobey, A. J. and A. M. Suggs (1963). *Control of Aircraft and Missile Powerplants*. John Wiley, New York.

APPENDIX

Numerical data

The integrated system is of the form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{W}w; \quad \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} + \Gamma w,$$

with $\mathbf{x}$, $\mathbf{u}$, and $\mathbf{y}$ as defined by equations (2), (3) and (4), respectively, and $w = \mathrm{WB3}$—the compressor bleed flow which is modelled as a disturbance on the plant. The system matrices are as follows.

$$\left[\begin{array}{ccccccccccccc}
-5.91\times10^{-2} & 0 & 7.18\times10^{-2} & 0 & -2.28\times10^{-1} & 0 & 4.59\times10^{-5} & -3.19\times10 & 1.12\times10^{-3} & 3.35\times10^{-4} & 6.92\times10^{-4} & 1.56\times10^{-5} & 1.45\times10^{-1} \\
0 & -1.50\times10^{-1} & 0 & 2.42\times10 & 0 & -1.32\times10^{2} & 3.19\times10 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1.37\times10^{-1} & 0 & -4.03\times10^{-1} & 0 & 1.30\times10 & 0 & 5.63\times10^{-6} & -3.92 & -7.21\times10^{-3} & -2.11\times10^{-3} & -4.36\times10^{-3} & -9.95\times10^{-5} & 1.55\times10^{-1} \\
0 & -2.14\times10^{-1} & 0 & -2.0 & 0 & 5.05\times10^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1.24\times10^{-2} & 0 & 1.89\times10^{-2} & 0 & -5.48\times10^{-1} & 0 & 0 & -1.40\times10^{-7} & 2.57\times10^{-4} & 7.68\times10^{-5} & 1.60\times10^{-4} & 3.69\times10^{-6} & -3.30\times10^{-3} \\
0 & 3.17\times10^{-3} & 0 & -4.16\times10^{-2} & 0 & -1.29\times10^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1.0 & 0 & 1.23\times10^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-7.29\times10^{-1} & 0 & -1.28\times10^{-1} & 0 & 0 & 0 & 0 & -1.26\times10^{-3} & -5.30 & 5.58 & 2.91 & -1.66\times10^{-2} & 5.05\times10 \\
3.56\times10^{-1} & 0 & 6.26\times10^{-2} & 0 & 0 & 0 & 0 & 6.14\times10^{-4} & 7.03\times10^{-1} & -4.36 & 5.76 & 7.50\times10^{-1} & -1.44 \\
4.85 & 0 & 8.54\times10^{-1} & 0 & 0 & 0 & 0 & 8.38\times10^{-3} & -1.95 & -1.53\times10 & -3.48\times10 & 2.88\times10 & 2.70\times10 \\
1.52\times10^{-1} & 0 & 2.68\times10^{-2} & 0 & 0 & 0 & 0 & 2.63\times10^{-4} & 3.42\times10^{-1} & 5.98 & 1.72 & -3.50\times10 & -2.98 \\
-4.62\times10^{-1} & 0 & -8.13\times10^{-2} & 0 & 0 & 0 & 0 & -7.98\times10^{-4} & 1.53 & 4.52\times10^{-1} & 9.39\times10^{-1} & 2.06\times10^{-2} & -1.99\times10^{2}
\end{array}\right]$$

$$
B = \left[\begin{array}{cccccc}
-3.18\times10^{-2} & 9.17\times10^{-7} & 9.01\times10^{-7} & -1.03\times10^{-1} & 6.16\times10^{-1} & -8.50\times10^{-6} \\
0 & 2.99\times10^{-2} & 3.85\times10^{-2} & 0 & 0 & 0 \\
-2.13\times10^{-1} & 1.13\times10^{-7} & -7.34\times10^{-6} & -8.06\times10^{-2} & 6.59\times10^{-1} & 0 \\
0 & -2.96\times10^{-1} & 2.08\times10^{-2} & 0 & -5.15\times10^{-3} & 2.40 \\
-2.34\times10^{-2} & -2.61\times10^{-7} & 1.31\times10^{-7} & 2.63\times10^{-1} & -2.82\times10^{-2} & 0 \\
0 & 1.01\times10^{-3} & -9.94\times10^{-3} & 0 & 1.98\times10^{-1} & -8.63\times10^{-3} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right.
$$

$$
\left.\begin{array}{cccccc}
5.86\times10^{-6} & 4.26\times10^{-2} & -7.99\times10^{-2} & 8.81\times10^{-3} & -1.97\times10^{-1} & -1.42\times10^{-5} \\
0 & 0 & 0 & 0 & 0 & 0 \\
-6.83\times10^{-5} & 7.32\times10^{-2} & -1.10\times10^{-1} & 3.01\times10^{-2} & -7.84\times10^{-2} & -8.86\times10^{-2} \\
0 & 0 & 0 & 0 & 0 & 0 \\
3.52\times10^{-6} & -1.89\times10^{-3} & 2.99\times10^{-2} & -4.39\times10^{-3} & -1.71\times10^{-2} & -2.20\times10^{-2} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
3.13\times10^{-2} & 2.38\times10 & 4.30\times10 & 2.24\times10 & -6.63\times10^{-5} & 1.73\times10^{-2} \\
4.04\times10^{-2} & -7.59\times10^{-1} & -1.35 & -6.81\times10^{-1} & -2.98\times10^{-3} & 4.21\times10^{-2} \\
4.78 & 1.26\times10 & 2.29\times10 & 1.19\times10 & 1.22\times10^{2} & -1.06\times10^{-2} \\
2.94\times10^{-2} & -1.47 & -2.49 & -1.30 & -5.02\times10^{-3} & -1.57\times10^{-3} \\
1.73\times10^{-2} & -1.57\times10 & -2.84\times10 & -1.48\times10 & -5.08\times10^{-3} & -7.51\times10^{-3}
\end{array}\right]
$$

$$
W = [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; -8.59\times10 \;\; -1.28\times10^{2} \;\; 3.78\times10^{2} \;\; -8.18\times10 \;\; -3.18\times10]^{T},
$$

$$
C = \left[\begin{array}{ccccccc}
9.85\times10^{-1} & 0 & 1.73\times10^{-1} & 0 & 0 & 0 & 0 \\
-8.19\times10^{-2} & 0 & 8.84\times10^{-4} & 0 & 7.37\times10^{-5} & 0 & 4.61\times10^{-5} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 5.73\times10 & 0 & 0 \\
7.35\times10^{-2} & 0 & -4.18\times10^{-1} & 0 & 0 & 0 & 7.15\times10^{-5} \\
0 & 0 & 0 & 0 & 0 & 0 & 5.73\times10 \\
0 & 0 & 0 & 5.73\times10 & 0 & 0 & 0 \\
0 & 4.24\times10^{-1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 5.73\times10 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -6.35\times10^{-2} & 0 & 1.03\times10 & 0 & -5.61\times10 & 1.35\times10
\end{array}\right.
$$

$$
\left.\begin{array}{cccccc}
1.53\times10^{-3} & -5.08\times10^{-6} & -1.48\times10^{-6} & -3.05\times10^{-6} & -6.92\times10^{-8} & 1.09\times10^{-4} \\
-3.21\times10 & -1.46\times10^{-4} & -3.57\times10^{-5} & -7.45\times10^{-5} & -1.84\times10^{-6} & 1.69\times10^{-1} \\
5.73\times10 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
5.73\times10 & -8.64\times10^{-8} & -2.49\times10^{-8} & -5.13\times10^{-8} & -1.16\times10^{-9} & 1.85\times10^{-6} \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
$$

$$
\begin{bmatrix}
0 & 0 & 0 & -1.09\times10^{-4} & 1.74\times10^{-4} & 0 & -4.06\times10^{-8} & 5.18\times10^{-5} & 9.33\times10^{-5} & -1.76\times10^{-6} & 8.04\times10^{-5} & 3.45\times10^{-8} \\
-6.82\times10^{-2} & 9.23\times10^{-7} & -3.84\times10^{-7} & -1.15\times10^{-1} & 7.21\times10^{-1} & 0 & -6.06\times10^{-6} & 5.46\times10^{-2} & -9.78\times10^{-2} & 1.39\times10^{-2} & -2.08\times10^{-1} & -1.54\times10^{-2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2.86\times10^{-6} & -2.86\times10^{-6} & -2.86\times10^{-6} & 5.54\times10^{-6} & -4.19\times10^{-5} & 8.68\times10^{-6} & 0 & 8.77\times10^{-7} & 4.06\times10^{-6} & -3.57\times10^{-7} & 1.88\times10^{-6} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1.27\times10^{-2} & 1.63\times10^{-2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$$\Gamma = [0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0\quad 0]^T.$$

The scaling used to normalize the controlled variables ($\mathbf{z}$) and the control inputs ($\mathbf{u}$) are as follows:

$$\mathbf{z}_{cmax} = [7.6\quad 6.3\quad 4.0\quad 10.2\quad 5.0\quad 200.0]^T,$$
$$\mathbf{u}_{max} = [5.0\quad 5.0\quad 10.0\quad 0.7\quad 0.7\quad 0.7\quad 1000.0\quad 20.0\quad 8.0\quad 45.0\quad 10.0\quad 10.0]^T.$$

The sensitivity weightings W_{S_i} and the complementary sensitivity weightings W_{T_i} for each of the controlled variables z_i was chosen to be of the form:

$$W_{S_i} = \frac{a_i s + 1000}{10 a_i s + 1}, \quad W_{T_i} = \frac{b_i s}{0.001 b_i s + 1},$$

with $a_i = 27.92,\ 11.17,\ 31.41,\ 22.33,\ 22.33,\ 13.4$, and $b_i = 0.233,\ 0.093,\ 0.263,\ 0.185,\ 0.185,\ 0.111$ for $i = 1, \ldots, 6$.
The maximum values for the control rates ($\dot{\mathbf{u}}_{max}$) used to determine $\mathbf{W}_{\dot{u}}$ were chosen to be:

$$\dot{\mathbf{u}}_{max} = [52\quad 52\quad 120\quad 3\quad 6\quad 3\quad 36{,}000\quad 240\quad 45\quad 240\quad 20\quad 27.5]^T.$$

The first order actuators used in the H_∞ control design are of the form

$$\frac{\quad}{u_{i_c}} = \frac{\quad}{s + a_i},$$

with $a_i = 20,\ 20,\ 20,\ 6,\ 10,\ 6,\ 13,\ 20,\ 10,\ 20,\ 5,\ 6$ for $i = 1, \ldots, 12$.
The 14th order controller, obtained after reduction of the H_∞ optimal controller is of the form:

$$\dot{\mathbf{x}}_c = \mathbf{F}\mathbf{x}_c + \mathbf{G}\begin{bmatrix} z_c - z \\ y \\ \mathrm{WB3} \end{bmatrix}; \quad \mathbf{u}_c = \mathbf{H}\mathbf{x}_c + \mathbf{J}\begin{bmatrix} z_c - z \\ y \\ \mathrm{WB3} \end{bmatrix},$$

with the controller state-space matrices as follows:

$$\mathbf{F} = \text{Block diagonal}\Big[-3.19\times10^{-3},\ -3.54\times10^{-3},\ -4.48\times10^{-3},\ -4.48\times10^{-3},\ -7.44\times10^{-3},\ -8.95\times10^{-3}$$
$$-1.82\times10^{-1},\ -8.37\times10^{-1},\ \begin{pmatrix} -1.35 & 7.49\times10^{-2} \\ -7.49\times10^{-2} & -1.35 \end{pmatrix},\ -2.36,\ -3.02,\ \begin{pmatrix} -7.75 & 5.63 \\ -5.63 & -7.75 \end{pmatrix}\Big],$$

$$
\mathbf{G} = \left[\begin{matrix}
-3.91\times10^{-1} & -3.52\times10^{-2} & -1.14\times10 & -4.70\times10^{3} & 2.51\times10^{-2} & 3.04\times10^{-3} \\
1.78\times10 & 1.67\times10^{1} & 3.07\times10^{-1} & 2.93\times10^{-2} & -1.56\times10^{-1} & -2.05\times10^{-2} \\
2.47\times10^{-2} & -8.15\times10^{3} & -4.45\times10^{-3} & -4.85\times10^{-1} & 2.59 & -5.74\times10^{6} \\
1.83\times10^{-3} & -1.92\times10^{3} & -1.81\times10^{-3} & -6.12\times10^{1} & -1.18 & 5.73\times10^{-5} \\
-2.06\times10^{-2} & 7.45\times10^{2} & 8.55\times10^{-4} & -1.03\times10^{-3} & 5.58\times10^{-3} & -1.20 \\
1.88\times10^{-3} & -1.46\times10 & 1.03\times10^{4} & 1.03\times10^{-4} & -5.16\times10^{4} & -7.24\times10^{-4} \\
1.73\times10 & 2.11\times10 & -3.02 & 6.58\times10^{1} & -3.50 & -1.53\times10^{-1} \\
4.66\times10^{-1} & -8.42\times10^{-1} & -1.30 & 2.21\times10^{-2} & -1.43\times10^{1} & -1.35\times10^{-2} \\
1.30 & 1.20\times10 & -1.07\times10 & 1.36\times10^{1} & -7.07\times10^{-1} & -7.79\times10^{-2} \\
-1.10 & -4.34 & 3.05 & -7.10\times10^{2} & 3.00\times10^{-1} & -2.79\times10^{2} \\
-3.62\times10^{1} & -1.51 & -2.21 & 2.27\times10^{2} & -7.31\times10^{2} & -1.26\times10^{-2} \\
3.38\times10^{1} & 2.17 & -3.48 & -6.24\times10^{-3} & -3.97\times10^{2} & 5.08\times10^{2} \\
-1.65\times10^{4} & 2.26 & -9.67\times10^{2} & 2.31\times10^{2} & -4.47\times10^{-2} & 1.14\times10^{2} \\
5.44\times10^{-1} & -7.20 & 1.30\times10^{1} & -1.26\times10^{2} & -8.43\times10^{4} & -2.50\times10^{2}
\end{matrix}\right.
$$

$$
\begin{matrix}
3.61\times10^{5} & -1.04\times10^{5} & 3.83\times10^{6} & 8.03\times10^{6} & 2.67\times10^{5} & 1.84\times10^{-4} \\
-1.14\times10^{4} & 9.83\times10^{5} & -1.40\times10^{-4} & -1.70\times10^{4} & -2.04\times10^{-4} & -3.39\times10^{4} \\
-2.09\times10^{7} & -3.96\times10^{6} & 1.59\times10^{5} & 1.85\times10^{5} & 1.76\times10^{6} & -1.77\times10^{-4} \\
-3.17\times10^{-6} & -6.49\times10^{-6} & 1.07\times10^{5} & 1.50\times10^{-5} & 3.23\times10^{6} & 6.34\times10^{-4} \\
-1.68\times10^{-5} & 2.18\times10^{6} & -4.61\times10^{6} & 4.13\times10^{-6} & -7.67\times10^{5} & -3.16\times10^{-5} \\
-4.75\times10^{-6} & -1.47\times10^{5} & 1.95\times10^{5} & 3.06\times10^{5} & 4.98\times10^{-6} & 5.83\times10^{5} \\
-2.99\times10^{1} & -3.87\times10^{1} & 7.69\times10^{1} & 9.96\times10^{-1} & 1.70\times10^{1} & -4.96\times10^{2} \\
1.13 & 1.47 & -2.92 & -3.78 & -6.46\times10^{1} & 7.26\times10^{-2} \\
-1.88 & -2.44 & 4.85 & 6.28 & 1.07 & -2.28\times10^{-2} \\
5.18\times10^{-1} & 6.67\times10^{1} & -1.34 & -1.73 & -2.95\times10^{1} & -9.45\times10^{-1} \\
-3.53\times10^{1} & -4.58\times10^{1} & 9.11\times10^{1} & 1.18 & 2.01\times10^{-1} & 6.15\times10^{3} \\
-7.28\times10^{1} & -9.41\times10^{1} & 1.88 & 2.43 & 4.15\times10^{1} & 3.43\times10^{2} \\
4.36\times10^{1} & 5.65\times10^{1} & -1.12 & -1.45 & -2.48\times10^{1} & -2.61\times10^{-3} \\
-2.06 & -2.67 & 5.31 & 6.87 & 1.17 & 5.83\times10^{-4}
\end{matrix}
$$

$$
\left.\begin{matrix}
1.65\times10^{-4} & -1.14\times10^{4} & 3.46\times10^{5} & 4.64\times10^{5} & 5.63\times10^{5} & 6.80\times10^{-4} \\
-3.05\times10^{4} & 2.37\times10^{4} & -6.77\times10^{5} & -3.00\times10^{5} & -1.04\times10^{4} & -1.17\times10^{3} \\
-1.68\times10^{4} & 2.89\times10^{-5} & 1.69\times10^{5} & -4.29\times10^{6} & -9.29\times10^{5} & -6.43\times10^{4} \\
2.35\times10^{4} & 1.68\times10^{3} & 2.02\times10^{4} & 2.37\times10^{5} & -2.57\times10^{1} & 1.44\times10^{4} \\
-2.90\times10^{5} & -3.85\times10^{4} & 1.21\times10^{-4} & -3.59\times10^{5} & -7.39\times10^{-5} & 5.31\times10^{-4} \\
5.17\times10^{5} & -2.50\times10^{-4} & 7.55\times10^{5} & -4.42\times10^{5} & -1.34\times10^{5} & 5.99\times10^{4} \\
-4.12\times10^{2} & -6.25\times10^{1} & 1.78\times10^{1} & -1.49\times10^{2} & -9.95\times10^{2} & -2.72\times10^{2} \\
6.94\times10^{2} & -9.25\times10^{2} & 4.22\times10^{-2} & -7.45\times10^{3} & 2.20\times10^{-3} & 7.19\times10^{2} \\
-1.23\times10^{2} & -9.93\times10^{1} & 3.01\times10^{1} & -2.66\times10^{3} & -1.53\times10^{1} & 1.23\times10^{1} \\
-9.67\times10^{-1} & 6.36\times10^{1} & -5.37\times10^{1} & 8.28\times10^{3} & 3.34\times10^{2} & 7.92\times10^{-1} \\
1.52\times10^{2} & -2.75\times10^{1} & 8.17\times10^{2} & -2.07\times10^{2} & -3.46\times10^{2} & 1.94\times10^{1} \\
2.53\times10^{2} & -1.80\times10^{3} & 1.19\times10^{2} & 9.74\times10^{3} & 1.24\times10^{3} & -8.17\times10^{1} \\
3.49\times10^{2} & -7.77\times10^{1} & 2.88\times10^{-1} & 1.32\times10^{1} & -1.60\times10^{1} & -5.23\times10^{1} \\
-8.93\times10^{3} & -3.41\times10^{1} & 2.09\times10^{2} & 8.30\times10^{2} & 2.92\times10^{2} & 3.86\times10^{-1}
\end{matrix}\right]
$$

$$
\mathbf{H} = \left[\begin{matrix}
-1.03\times10^{1} & -1.32\times10^{1} & 1.61\times10^{1} & -5.06\times10^{3} & -4.5\times10^{3} & 1.87\times10^{1} & 1.50\times10^{1} \\
3.30\times10^{-4} & 7.63\times10^{-4} & 2.70\times10^{1} & 7.64\times10^{1} & 9.84\times10^{5} & -1.67\times10^{3} & -7.91\times10^{4} \\
2.54\times10^{2} & 4.69\times10^{-2} & 1.46 & -1.35\times10^{1} & 3.16\times10^{3} & -3.87\times10^{-2} & -2.99\times10^{-2} \\
-1.12\times10^{3} & -1.72\times10^{3} & -9.29\times10^{4} & -5.83\times10^{4} & -5.31\times10^{4} & -7.76\times10^{4} & -6.74\times10^{-5} \\
1.74\times10^{3} & 3.03\times10^{3} & -1.23\times10^{1} & -1.62\times10^{2} & 6.18\times10^{4} & -1.10\times10^{3} & -1.73\times10^{3} \\
-6.28\times10^{5} & -1.47\times10^{4} & -1.75\times10^{2} & -6.92\times10^{2} & -1.57\times10^{5} & 2.03\times10^{-4} & 1.26\times10^{-4} \\
-1.89\times10 & -2.10\times10 & 2.63\times10 & -7.98\times10^{1} & -1.41\times10 & 2.25\times10 & 2.48\times10 \\
4.04\times10^{1} & 5.96\times10^{1} & -4.98\times10^{1} & 4.36\times10^{2} & -6.71\times10^{2} & -5.21\times10^{1} & -5.21\times10^{-1} \\
-1.32\times10^{1} & -2.19\times10^{1} & 1.22\times10^{1} & -2.71\times10^{-2} & -4.73\times10^{2} & 1.32\times10^{1} & 1.48\times10^{-1} \\
7.98\times10^{1} & 9.97\times10^{1} & -1.17 & 4.54\times10^{2} & -3.25\times10^{1} & -9.84\times10^{-1} & -1.16 \\
3.43\times10^{-2} & -8.45\times10^{2} & -1.24\times10^{1} & -3.42\times10^{2} & -1.12\times10^{2} & -4.39\times10^{-2} & -9.34\times10^{-2} \\
-1.34\times10^{-1} & -2.20\times10^{1} & 2.96\times10^{1} & 1.50\times10^{-5} & -3.63\times10^{3} & 2.59\times10^{-1} & 2.83\times10^{-1}
\end{matrix}\right.
$$

$$
\left.\begin{matrix}
9.79\times10^{-2} & -1.86\times10^{1} & 4.70\times10^{3} & 3.66\times10^{1} & -5.76\times10^{-2} & 5.67\times10^{2} & 1.33\times10^{1} \\
-1.15\times10^{3} & 2.71\times10^{-2} & 6.30\times10^{2} & -2.22\times10^{-1} & 1.43\times10^{1} & -4.47\times10^{-2} & 3.89\times10^{-3} \\
-4.89\times10^{2} & 2.09\times10^{1} & 3.39\times10^{1} & -8.75\times10^{1} & 1.70\times10^{-1} & -9.92\times10^{-2} & 3.95\times10^{2} \\
7.81\times10^{-3} & 3.71\times10^{3} & 1.04\times10^{3} & -1.67\times10^{-2} & 1.53\times10^{-2} & 8.85\times10^{-3} & -2.66\times10^{2} \\
-3.65\times10^{3} & -1.07\times10^{2} & -2.76\times10^{-2} & 3.46\times10^{-2} & -1.36\times10^{2} & -3.31\times10^{-3} & 8.12\times10^{-3} \\
2.44\times10^{4} & -2.12\times10^{4} & -6.34\times10^{-4} & -2.36\times10^{-4} & -6.87\times10^{4} & -6.52\times10^{4} & 4.21\times10^{-4} \\
1.70\times10 & -2.31\times10 & 7.52\times10^{1} & 1.00\times10 & 1.29\times10^{-1} & 2.23\times10 & -1.17\times10 \\
-5.65\times10^{1} & 5.41\times10^{1} & -2.53\times10^{2} & -7.61\times10^{1} & -1.54\times10^{-1} & 1.18\times10^{-1} & 3.20\times10^{-1} \\
3.25\times10^{1} & -9.06\times10^{2} & 3.82\times10^{-2} & -1.32\times10^{-1} & 4.60\times10^{-1} & -1.15\times10^{1} & -4.29\times10^{-1} \\
-1.19 & 1.27 & -1.07\times10^{-1} & -9.12\times10^{1} & -2.42 & 1.53 & 1.40 \\
1.60\times10^{-1} & 1.77\times10^{1} & 1.11\times10^{2} & -9.90\times10^{1} & 2.65\times10^{-1} & -6.82\times10^{-2} & 2.46\times10^{-1} \\
9.97\times10^{-2} & -4.66\times10^{1} & 3.60\times10^{-2} & 1.47 & 4.47\times10^{1} & -2.13\times10^{-2} & 1.16\times10^{-1}
\end{matrix}\right]
$$

$$
\mathbf{J}=\left[\begin{matrix}
7.57\times10^{-4} & -6.80\times10^{-3} & -5.96\times10^{-2} & 2.23\times10^{4} & 4.04\times10^{-4} & 1.21\times10^{3}\\
9.09\times10^{-4} & 7.83\times10^{-4} & 1.71\times10^{2} & -7.76\times10^{-3} & -6.49\times10^{-3} & -4.00\times10^{-4}\\
-3.43\times10^{-2} & 2.69\times10^{-2} & -4.91\times10^{-2} & 1.93\times10^{2} & -2.56\times10^{-2} & 2.17\times10^{-3}\\
-1.02\times10^{-4} & 4.61\times10^{-4} & 3.79\times10^{-3} & 1.19\times10^{6} & -6.89\times10^{-5} & -7.86\times10^{5}\\
-2.36\times10^{-3} & 2.03\times10^{-3} & -3.20\times10^{3} & 8.32\times10^{-4} & -2.23\times10^{-3} & 1.80\times10^{-4}\\
-4.06\times10^{-4} & 3.31\times10^{-4} & -6.35\times10^{-4} & 2.44\times10^{-3} & 1.96\times10^{-3} & 2.85\times10^{-5}\\
1.28 & 7.76 & 2.13\times10 & 1.03\times10^{-1} & -3.03\times10^{1} & 1.24\\
1.09\times10^{-3} & 1.91\times10^{-2} & 1.02\times10^{-1} & -2.92\times10^{-5} & -9.79\times10^{-4} & 3.50\times10^{-5}\\
-1.38\times10^{-3} & 7.47\times10^{-3} & 5.73\times10^{2} & -2.49\times10^{5} & -9.81\times10^{-4} & -9.81\times10^{4}\\
-2.16\times10^{-3} & 3.71\times10^{-2} & 2.92\times10^{-1} & -1.16\times10^{3} & -1.54\times10^{-3} & -4.62\times10^{-3}\\
-1.84\times10^{-3} & 1.24\times10^{-3} & 1.08\times10^{-2} & 4.19\times10^{4} & -1.13\times10^{3} & -4.83\times10^{-4}\\
2.04\times10^{-4} & 1.12\times10^{-3} & 1.49\times10^{2} & -2.43\times10^{4} & 3.02\times10^{-4} & -3.61\times10^{-4}
\end{matrix}\right.
$$

$$
\begin{matrix}
-1.93\times10^{-2} & -2.52\times10^{2} & 4.98\times10^{-2} & 6.48\times10^{-2} & 1.10\times10^{2} & -2.20\times10^{-4}\\
5.11\times10^{-3} & 6.62\times10^{3} & -1.32\times10^{-2} & -1.71\times10^{2} & -2.91\times10^{-5} & 2.09\times10^{-4}\\
-3.72\times10^{-3} & -4.71\times10^{3} & 9.58\times10^{-3} & 1.23\times10^{-2} & 2.12\times10^{3} & 2.53\times10^{-3}\\
1.24\times10^{-3} & 1.61\times10^{3} & -3.21\times10^{-3} & -4.16\times10^{3} & -7.08\times10^{4} & 1.75\times10^{5}\\
-1.68\times10^{-4} & -1.73\times10^{4} & 4.33\times10^{4} & 5.15\times10^{4} & 9.57\times10^{-5} & 1.98\times10^{-4}\\
-5.27\times10^{-5} & -6.68\times10^{5} & 1.36\times10^{4} & 1.74\times10^{4} & 3.00\times10^{-5} & -6.88\times10^{5}\\
7.41 & 9.52 & -1.91\times10 & -2.47\times10 & -4.22 & 5.84\times10^{-2}\\
3.40\times10^{2} & 4.55\times10^{2} & -8.78\times10^{-2} & -1.16\times10^{1} & -1.94\times10^{-2} & 3.97\times10^{4}\\
1.88\times10^{-2} & 2.39\times10^{2} & -4.86\times10^{-2} & -6.28\times10^{2} & -1.07\times10^{-2} & 2.48\times10^{-4}\\
9.52\times10^{2} & 1.23\times10^{-1} & -2.46\times10^{1} & -3.19\times10^{-1} & -5.43\times10^{2} & 1.02\times10^{3}\\
3.80\times10^{-3} & 4.43\times10^{3} & -9.79\times10^{3} & -1.23\times10^{2} & -2.16\times10^{3} & 1.29\times10^{4}\\
4.65\times10^{-3} & 6.02\times10^{3} & -1.20\times10^{2} & -1.55\times10^{2} & -2.65\times10^{3} & 1.78\times10^{-5}
\end{matrix}
$$

$$
\left.\begin{matrix}
-4.70\times10^{4} & -1.28\times10^{2} & -2.32\times10^{3} & 1.21\times10^{2} & 4.06\times10^{3} & -6.83\times10^{2}\\
1.63\times10^{-1} & -8.35\times10^{1} & 7.80\times10^{2} & -3.13\times10^{3} & 7.80\times10^{2} & 1.38\times10^{2}\\
4.68\times10^{2} & -3.90 & 1.38 & -4.34\times10^{3} & -7.60\times10^{-1} & -1.31\times10^{2}\\
-1.79\times10^{-4} & 6.04\times10^{3} & -2.27\times10^{3} & -9.37\times10^{4} & 1.29\times10^{3} & 4.11\times10^{-3}\\
-6.66\times10^{3} & 3.19\times10^{1} & -1.16\times10^{1} & -2.65\times10^{4} & 6.47\times10^{2} & -4.22\times10^{5}\\
-1.19\times10^{-2} & 8.63\times10^{2} & -1.36\times10^{2} & -6.86\times10^{5} & -1.82\times10^{3} & -2.05\times10^{4}\\
-1.91\times10^{-1} & 4.88 & -2.22 & -3.42 & 1.42 & 4.40\times10\\
1.07\times10^{2} & -2.34\times10^{1} & 1.15\times10^{1} & -3.27\times10^{2} & -7.58\times10^{2} & 5.31\times10^{-1}\\
-5.83\times10^{3} & 1.49\times10^{1} & -6.40\times10^{2} & -2.10\times10^{-2} & 3.94\times10^{2} & 1.06\times10^{1}\\
6.42\times10^{3} & -1.12\times10^{1} & 7.96\times10^{2} & -9.87\times10^{2} & -5.98\times10^{-2} & 8.44\times10^{1}\\
-4.40\times10^{3} & 3.84\times10^{2} & -3.38\times10^{2} & -5.26\times10^{3} & 2.65\times10^{-2} & -3.47\times10^{3}\\
1.19\times10^{4} & -9.65\times10^{3} & 5.40\times10^{3} & -2.52\times10^{3} & -3.81\times10^{3} & 1.90\times10^{3}
\end{matrix}\right].
$$

The compressor bleed flow due to RCS area commands was calculated as follows for the controller evaluation:

$$\mathrm{WB3} = 4.41 \times |\mathrm{AQR}| + 4.41 \times |\mathrm{AYR}| + 8.82 \times |\mathrm{ARR}|.$$

The output matrices for the gross thrusts from the three "nozzles", corresponding to the results presented under the controller evaluation section are given by:

$$\begin{bmatrix} \mathrm{FG9} \\ \mathrm{FGE} \\ \mathrm{FGV} \end{bmatrix} = \mathbf{C}_{FG}\mathbf{x} + \mathbf{D}_{FG}\mathbf{u} + \Gamma_{FG} w,$$

with

$$
\mathbf{C}_{FG}=\begin{bmatrix}
2.20\times10^{-4} & 0 & 3.87\times10^{5} & 0 & 0 & 0 & 0 & -3.79\times10^{7} & 1.81\times10^{3} & -1.99\times10^{3} & -9.63\times10^{3} & -1.71\times10^{4}\\
-5.53\times10^{-1} & 0 & -9.73\times10^{2} & 0 & 0 & 0 & 0 & -9.55\times10^{4} & 1.83 & 5.41\times10^{1} & 1.12 & 2.58\times10^{2}\\
-6.87\times10^{-1} & 0 & -1.21\times10^{1} & 0 & 0 & 0 & 0 & -1.19\times10^{3} & 2.28 & 6.64\times10^{1} & 1.37 & 3.11\times10^{-2}
\end{bmatrix},
$$

$$
\mathbf{D}_{FG}=\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & -4.25\times10^{-4} & 3.05\times10 & 4.52\times10^{-2} & 2.43\times10^{2} & 5.37\times10^{-3} & 1.70\times10^{3}\\
0 & 0 & 0 & 0 & 0 & 0 & 2.07\times10^{2} & -1.88\times10 & 1.05\times10^{2} & -1.76\times10 & 5.32\times10^{3} & -2.76\times10^{-2}\\
0 & 0 & 0 & 0 & 0 & 0 & 1.84\times10^{-2} & -2.33\times10 & -4.20\times10 & 4.36\times10^{1} & 4.89\times10 & -1.59\times10^{2}
\end{bmatrix},
$$

$$\Gamma_{FG} = [6.59\times10^{-2} \quad -3.80\times10 \quad -4.74\times10]^{T}.$$

The total gross thrust is given by:

$$\mathrm{FGT} = 1.027 \times \mathrm{FG9} + 0.453 \times \mathrm{FGE} + 1.194 \times \mathrm{FG9}.$$

The K_{ψ}^{ij} used for robustness analysis were as follows:

i, j:	1, 5	2, 4	2, 6	3, 1	3, 3	3, 5	4, 4	4, 6	5, 1	5, 5	6, 2	6, 4	6, 6.
$K_{\psi}^{ij} \times 10^{-2}$:	1.25	1.25	1.25	1.71	1.25	1.25	1.25	1.25	1.95	1.25	3.60	1.25	1.25.

Automatica, Vol. 29, No. 1, pp. 147–156, 1993
Printed in Great Britain.

0005–1098/93 $6.00 + 0.00

Robust Performances Control Design for a High Accuracy Calibration Device*

M. MILANESE, G. FIORIO and S. MALAN†

A high accuracy calibration device has been modelled, and parametric and unmodeled dynamic errors have been evaluated. A controller was designed taking into account performance specifications, and robustness analysis was performed dealing with the derived nonlinear parametric and dynamic perturbations.

Key Words—Control applications; identification; error analysis; parameter estimation; robust control; robustness analysis.

Abstract—This paper presents a case study of robust performances control design. The physical plant under examination consists of a platform for calibration of high accuracy accelerometers, which has to assume the properties of an inertial body, despite the vibrations coming from the surrounding ground. Plant modelling and parameter estimation, control system design and robustness analysis of the designed controllers are described and discussed. Besides a simplified model of the plant, called the nominal model, perturbations are also considered, taking into account parametric and dynamic uncertainties. The procedure followed for estimating model parameters, based on an unknown but bounded approach, is illustrated, and uncertainty intervals of parameter estimates are provided. Bounds of unstructured uncertainty are also derived from results of simulations aiming to evaluate the main effects of the unmodeled dynamics.

The design has been carried on through iterative steps of "nominal" design and robustness analysis. The design has been performed through H_∞ synthesis, based on the nominal model and taking into account the main performance specifications required for the present case study, i.e. stability, disturbance attenuation and command power limitation. The robustness analysis has been performed using recent techniques able to deal with frequency domain specifications and with mixed nonlinear parametric and dynamic perturbation, as required in the present case study.

1. INTRODUCTION

THE PROBLEM ORIGINATED from a national laboratory, in the realization of a calibration device for high accuracy acceleration transducers. This device requires to work over a platform, whose conditions should be close to those of an inertial body. Unfortunately, the laboratory is located in the neighborhoods of heavy mechanical factories, whose undesired effects are to generate vibrations in the surrounding ground.

In these conditions, it is required to reduce the r.m.s. value of the platform perturbing acceleration in the ratio 1 : 100 approximately, in order to have sufficiently negligible calibration errors with respect to the accuracy guaranteed for the most sensitive transducers.

It comes out from a preliminary analysis that it is not convenient to use simply passive mechanical elements, such as springs and dampers, to solve the problem of reducing platform perturbing acceleration, because unfeasible parameter values should be required. On the contrary, the use of active elements, such as electromagnetic actuators driven in feedback or in mixed feedback–feedforward control schemes, leads to much more effective disturbance attenuation.

Figure 1 gives a sketch of the plant. A concrete rectangular platform P is supported at each corner by a set of three elements lying on ground: a spring, a damper and an electromagnetic force generator. Another platform B, bearing the calibration device, is leant on the first one through similar mechanical elements, but without active generators.

Accuracy is the main concern of the calibration device, and the desired accuracy must be guaranteed for the controlled system, despite of the limited and uncertain information available on the plant. Then a robust (worst-case) approach was taken, based on the typical two steps.

- Identification of the approximate behaviour of the system in terms of a *nominal* model and a *perturbation* model, able to capture the discrepancies between the nominal model and the actual system (described in Sections 2 and 3).
- Design of a control system which assures acceptable performances according to some

* Received 30 August 1991; received in final form 17 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor J. Ackermann.

† Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy.

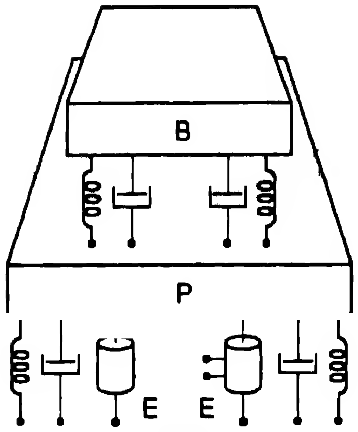


FIG. 1 The plant

given specifications, not only for the nominal model, but for all considered perturbations (described in Section 4).

It is well known that the type of perturbation model plays a key role. In the robust control literature, two main types of perturbations have been extensively studied: parametric (real) and dynamic (complex) perturbations, the former accounting for parametric variations, and the latter for unmodeled dynamics (see e.g Dorato and Yedavally, 1990 and Milanese *et al.*, 1989, for a collection of recent papers).

Recently, some work has been done on mixed types of perturbation (see e.g. Fan *et al.*, 1991, Milanese *et al.*, 1991), which seems able to better capture the physical information on the approximations introduced by the model. In fact a preliminary analysis of this case study has been performed in Fiorio *et al.* (1990) taking into account parametric perturbation only. The results clearly showed that in this case the unmodeled dynamics play a key role in robust stability. In this paper, we perform a more complete study, using a mixed perturbation model.

2 THE NOMINAL AND THE PERTURBATION MODELS

The aim of this section is to illustrate the results of model building for a suitable representation of the plant. First, the nominal model is described. Furthermore, perturbations on this model are introduced, taking into account its main approximations.

The nominal model is based on the following simplifying hypotheses. (1) The ground, as well as platforms P and B in Fig. 1, are considered as rigid bodies. (2) The ground has only vertical motion, and the structure has perfect symmetry at the four corners of the platforms. (3) Only linear equations are included in the model.

The four electromagnetic actuators at the corners of the lower platform of Fig. 1 are driven by the same electric current i, according to the

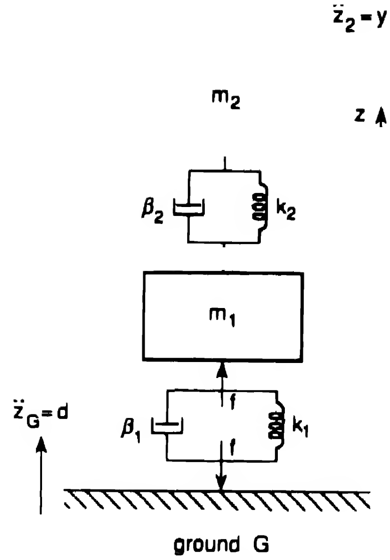


FIG 2 The simplified model

hypothesis that perfect symmetry gives rise to translation motion of the platforms along the vertical axis only. The simplified model is depicted in Fig. 2. In this figure, the four identical elements at each corner of the platforms are represented by only one equivalent parameter.

In the nominal model, the vertical component of the ground acceleration $\ddot{z}_G$ is considered as the only disturbance, which is denoted by d. The force f in Fig. 2 is considered as proportional to the current i, which is the only command variable of the system. So, it results: $f = K_F i$. The vertical component $\ddot{z}_2$ of the upper platform acceleration is the controlled output, denoted by y. Thus, the system can be represented by the SISO structure of Fig. 3, where $y(s)$, $i(s)$ and $d(s)$ represent system output, command variable and disturbance, respectively. $M(s)$ and $A(s)$ represent output sensor and actuator transfer functions, respectively. $C(s)$ is the regulator transfer function to be designed. $G(s)$ and $H(s)$ are the command to output and disturbance to output transfer functions, respectively. Their expressions are:

$$G(s) = N_G(S)D^{-1}(s), \quad (1)$$

$$H(s) = N_H(s)D^{-1}(s), \quad (2)$$

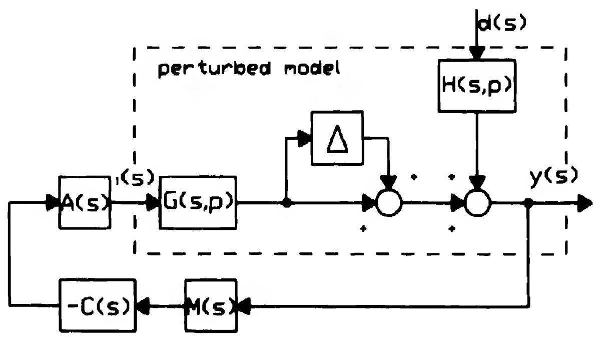


FIG. 3 The control scheme.

where

$$N_G(s) = K_F k_1^{-1} s^2 (1 + \beta_2 k_2^{-1} s), \quad (3)$$

$$N_H(s) = (1 + \beta_1 k_1^{-1} s)(1 + \beta_2 k_2^{-1} s), \quad (4)$$

$$\begin{aligned} D(s) = {} & m_1 m_2 k_1^{-1} k_2^{-1} s^4 \\ & + [m_1 \beta_2 + m_2(\beta_1 + \beta_2)] k_1^{-1} k_2^{-1} s^3 \\ & + [m_1 k_1^{-1} + m_2(k_1^{-1} + k_2^{-1}) \\ & + \beta_1 \beta_2 k_1^{-1} k_2^{-1}] s^2 \\ & + (\beta_1 k_1^{-1} + \beta_2 k_2^{-1}) s + 1. \end{aligned} \quad (5)$$

Output sensor is modelled by the transfer function

$$M(s) = K_m s (1 + s\tau)^{-} \quad (6)$$

Actuator transfer function $A(s)$ is assumed as a constant K_a in the frequency range of interest:

$$A(s) = K_a. \quad (7)$$

In order to take into account the approximations introduced by this simplified model, the three main assumptions are briefly discussed.

With regard to linearity assumption, the behaviour of the plant in normal operating conditions can be actually considered linear with good approximation, due to the very small displacements and accelerations present in these conditions, and to the fact that input current will be controlled to stay within the linearity range of the relation force–current.

On the contrary, while the rigid body assumption appears to be likely for platforms P and B, ground deformability gives parasitic effects which may cause stability problems in the closed loop. Then the deformability of the ground is taken into account by means of an uncertainty represented in terms of a multiplicative perturbation Δ as indicated in Fig. 3, such that

$$\|W^{-1}(\omega)\, \Delta(\omega)\|_\infty \le 1, \quad 0 \le \omega < \infty. \quad (8)$$

The choice of the weighting function $W(\omega)$ is discussed in Section 3.

The non perfect symmetry of the structure gives rise to rotating motion components of the platforms, with the effect of adding two pole–zero couples for each pole of the transfer functions of the nominal model. The pole and the zero of each added couple are very close, and close to the corresponding pole of the nominal model. Through simulation, we have verified that large asymmetries (up to 10% on each parameter) give transfer functions which can be recovered with very good approximation by suitable assessments of the parameters of the simplified model of Fig. 2.

In fact we assume that parameter vector p is known only to belong to a *parameter uncertainty set* Π. The identification procedure described in the next section provides parameter estimates with their ranges of variations, able to recover this source of parametric perturbations, as well as others, for example the ones due to errors in measurements.

3 IDENTIFICATION PROCEDURE

The model described in the previous section contains several parameters, whose values have to be known.

Constants K_m, τ, K_F and K_a are given the values from the data sheets of the corresponding components:

$$\begin{aligned} & K_m = 2 \times 10^5\ \mathrm{V\,sec^3\,m^{-1}}, \\ & \tau = 2\ \mathrm{sec}, \quad K_F = 8.7\ \mathrm{N\,A^{-1}}, \\ & K_a = 10.0\ \mathrm{AV^{-1}}. \end{aligned} \quad (9)$$

The remaining parameters of the model are: masses m_1 and m_2, stiffness coefficients k_1 and k_2, damping coefficients β_1 and β_2.

Also for these parameters, it could have been possible to disassemble the system and to measure them separately. Apart from practical difficulties, this approach is not appropriate because the assumed model is a simplified model, and its parameters have the nature of equivalent parameters, with implicit reference to some physical phenomena neglected in the model, as discussed in the previous section. For instance, stiffness parameter k_1 is an equivalent parameter, taking into account stiffness of the springs sustaining platform P, the ground elasticity, and asymmetry of the structure. From physical considerations, it can then be argued that parameter m_2 is the less affected one by neglected dynamics. Consequently, this parameter has been set to the value obtained by direct measurement:

$$m_2 = 440\ \mathrm{kg}. \quad (10)$$

The remaining parameters are identifiable from the given experimental conditions, and have been estimated from measurements on the overall system.

The available experimental data are:

- Samples of frequency response command to output (module and phase) $G(j\omega)$ with sinusoidal current generator supplying the actuator. Input and output measurements have been performed in open loop at about 20 frequencies in the range 0.5–40 Hz, and with some different command amplitudes: 0.5, 1, and 2 A r.m.s. values (Figs 4 and 5).
- Samples of frequency response disturbance to output $|H(j\omega)|$, (only module deduced

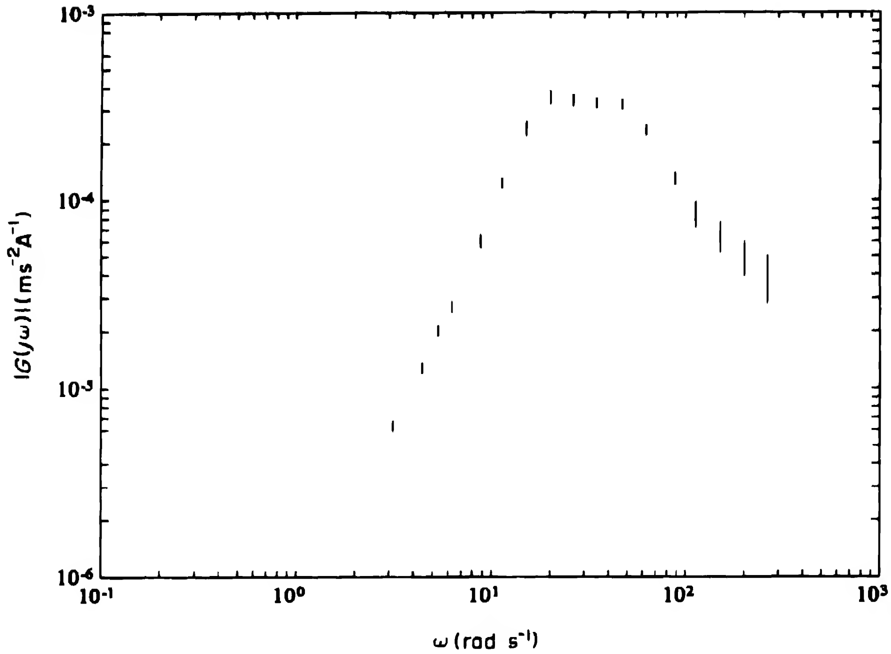


FIG. 4 Samples of $|G(j\omega)|$ with corresponding uncertainty intervals

from spectral densities of vertical acceleration of both ground and upper platform B in open loop normal operating conditions. Frequency range is 0.63–57 rad s^{-1}, with frequency resolution 1.25 rad s^{-1} (Fig. 6). This range of frequency contains more than 95% of both disturbance and output power (see Section 4).

These data can be described by the equation

$$y = F(p) + e, \tag{11}$$

where y is the vector containing the samples

$$y = [|G(j\omega_1)|, |G(j\omega_2)|, \dots, \arg G(j\omega_1), \arg G(j\omega_2), \dots, |H(j\omega_1)|, |H(j\omega_2)|, \dots], \tag{12}$$

p is the vector containing the unknown parameters

$$p = [k_1, k_2, \beta_1, \beta_2, m_1], \tag{13}$$

and e is an error term.

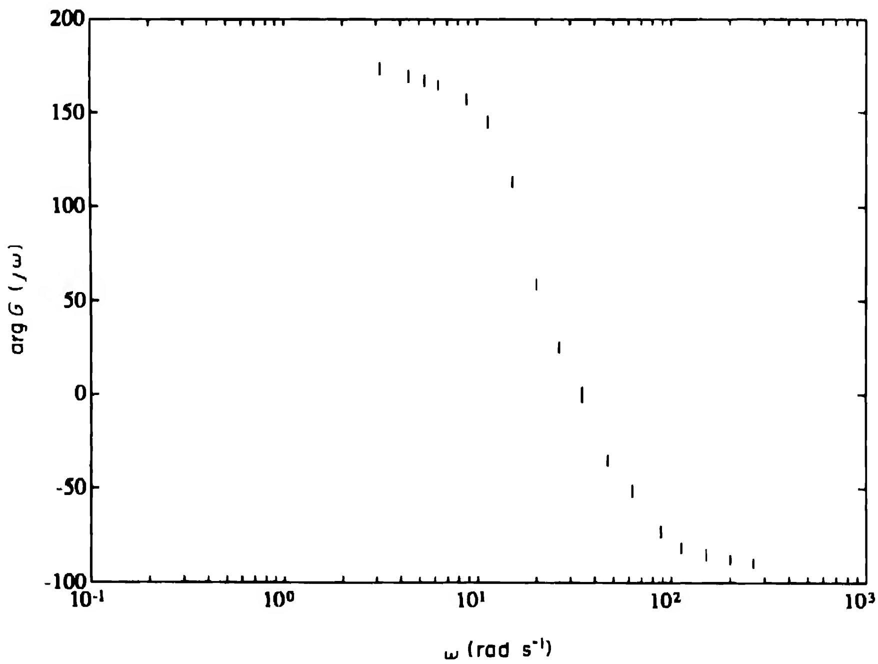


FIG 5 Samples of arg $G(j\omega)$ with corresponding uncertainty intervals

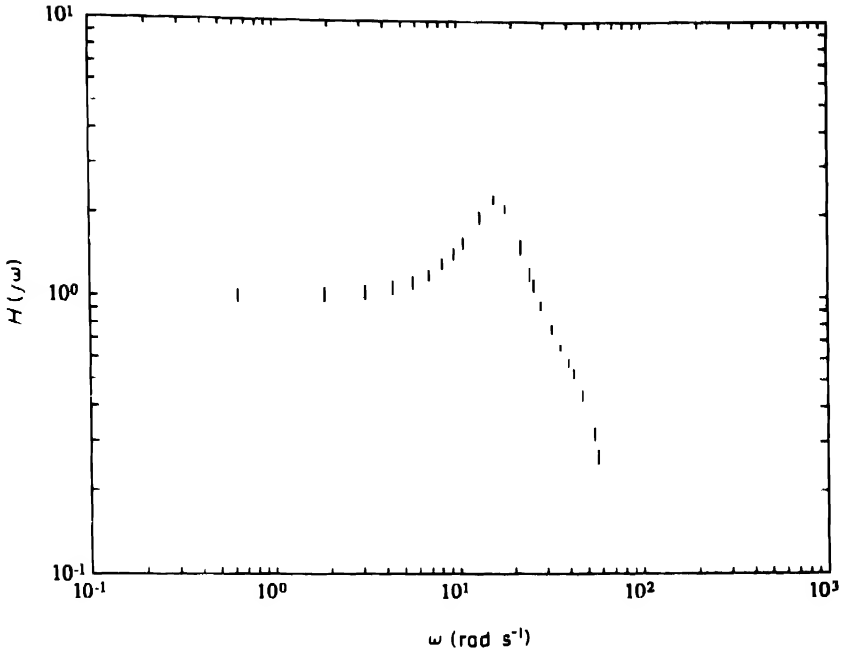


FIG 6. Samples of $|H(j\omega)|$ with corresponding uncertainty intervals.

Since accuracy is the main concern of the calibration problem, particular attention has been paid to the computation of parameter estimate accuracy, which depends on the knowledge about the error term e and on the estimation algorithm.

A statistical description of e does not appear appropriate, mainly due to the modelling errors discussed above. An unknown but bounded error approach has been adopted, which requires information on measurement error bounds only (see e.g. Schweppe, 1973; Milanese and Vicino, 1991b).

Uncertainty intervals v_i have been evaluated for each data, taking into account the accuracy of the measurement devices and the effects of unmodeled dynamics, expressed by the bounding function shown in Fig. 7. The used values are reported in Figs. 4–6. In such a way, the information on e is given as

$$\|e\|_\infty^v \le 1, \tag{14}$$

where $\|e\|_\infty^v = \max_i v_i^{-1}|e_i|$, $v_i > 0$ and v is the vector of error bounds corresponding to the uncertainty intervals on data.

A least square estimate p^0 of the unknown parameters is obtained as

$$p^0 := \arg\min_p \{\|y - F(p)\|_2^V\}, \tag{15}$$

where $\|\cdot\|_2^V$ denotes a weighted Euclidean norm, and V is a diagonal matrix with elements v_i.

Uncertainty in data induces uncertainty on the parameter estimated values. The maximal range of variation of the ith component p_i^0, due to possible errors consistent with (14) is indicated as estimate uncertainty interval u_i. The u_is have been evaluated by a method proposed by Belforte and Milanese (1981). The method is based on a quasilinearization technique and gives intervals certainly contained within the exact u_is. Indeed, it gives exact values if $\partial p_i^0/\partial y_j$ has constant sign for all $y \in Y$, the set Y being defined as

$$Y = \{y : \|y - \hat{y}\|_\infty^v \le 1\}. \tag{16}$$

where $\hat{y}$ denotes actual measurements.

This condition is difficult to be checked with certainty. However, $\partial p^0/\partial y$ has been evaluated at several grid points, providing good evidence that the condition is met. This has been accomplished by using the formula (see Belforte and Milanese, 1981)

$$\frac{\partial p^0}{\partial y} = (A^TVA)^{-1}A^TV, \tag{17}$$

where

$$A = \left.\frac{\partial F}{\partial p}\right|_{p^0(y)}. \tag{18}$$

In turn, $\dfrac{\partial F}{\partial p}$ has been computed as a function of p by means of the symbolic manipulation package DERIVE.

The identification procedure described above

has given the following results:

$$\begin{aligned} k_1 &= p_1 = p_1^0 \pm u_1 \\ &= (1.4 \pm 0.3)10^6\,\mathrm{N\,m^{-1}}, \\ k_2 &= p_2 = p_2^0 \pm u_2 \\ &= (1.0 \pm 0.3)10^6\,\mathrm{N\,m^{-1}}, \\ \beta_1 &= p_3 = p_3^0 \pm u_3 \\ &= (4.8 \pm 1.0)10^4\,\mathrm{N\,sec\,m^{-1}}, \\ \beta_2 &= p_4 = p_4^0 \pm u_4 \\ &= (1.7 \pm 0.3)10^4\,\mathrm{N\,sec\,m^{-1}}, \\ m_1 &= p_5 = p_5^0 \pm u_5 \\ &= (4.2 \pm 0.7)10^3\,\mathrm{kg}. \end{aligned} \tag{19}$$

Note that other estimators could give smaller estimate uncertainty intervals. In particular, methods to compute minimal uncertainty intervals estimators have been proposed by Milanese and Belforte (1982), Milanese and Tempo (1985) and Milanese and Vicino (1991a). However, a least square estimator has been adopted because it is computationally faster and has better robustness properties with respect to inexact knowledge of uncertainty error bounds v_i, as shown by Tempo and Wasilkowsky (1988).

In order to complete the perturbation model, the weighting function $W(\omega)$ of (8) have to be evaluated. To this extent effects of ground deformability has been simulated by space discretization of the ground. This discretization introduces several couples of zeros and poles in the transfer function $G(s)$ which are rather close to the imaginary axis. These low damped zero–pole couples give rise to peaks in the transfer function at relatively high frequencies ($\omega > 1000\,\mathrm{rad\,sec^{-1}}$, see Fig. 7). Now $W(\omega)$ have to be chosen so that $|G(j\omega, p^0)|\,(1 \pm |W(\omega)|)$ envelopes these peaks. The methods used for the design and the robustness analysis require $W(\omega)$ to be a rational stable function, whose order affects the complexity of the controller and the computational burden of the analysis. Then a first order, high pass (for $\omega > 1000\,\mathrm{rad\,sec^{-1}}$) function has been chosen:

$$W(\omega) = \frac{1.22j\omega}{j\omega + 1000}. \tag{20}$$

In summary, the perturbed model which has been adopted for the purpose of robust performances regulator design, is the mixed parametric and dynamically perturbed model represented in Fig. 3. The forms of $G(s, p)$ and $H(s, p)$ are given by equations (1)–(5). Symbols $G(s, p)$ and $H(s, p)$ denote explicitly the dependence of the plant transfer functions on the uncertain parameter vector p.

Parametric uncertainty is represented by the fact that parameter vector p is known only to belong to the *parameter uncertainty set* Π defined as

$$\Pi = \{p \in \mathbf{R}^5 : \|p - p^0\|_\infty^u \leq 1\}, \tag{21}$$

where p^0 denotes the 'nominal' parameter vector, whose components are given by the p_i^0's in equations (19), and u is the vector whose components are the estimates uncertainty intervals u_i in the same equations.

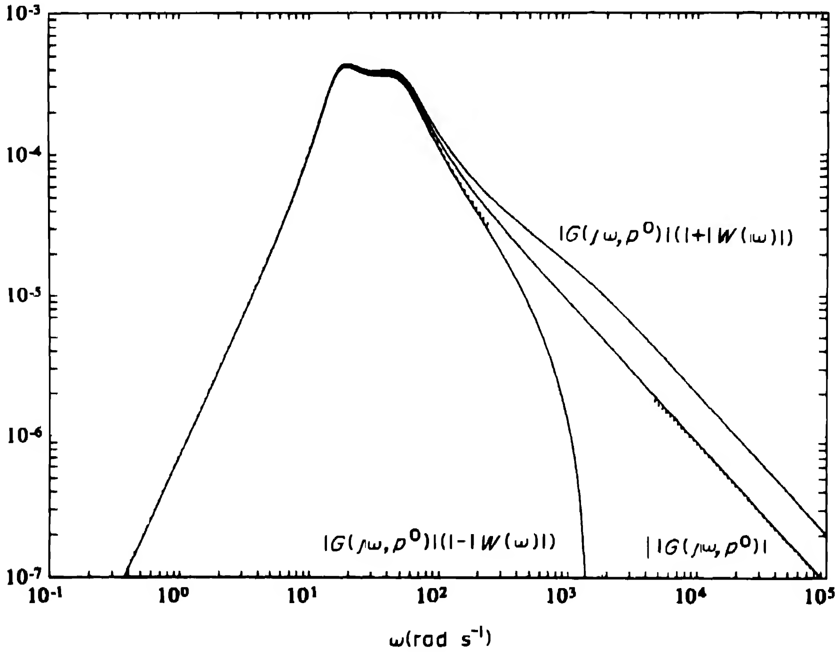


FIG 7 Bounds of perturbed transfer function $G(s, p^0)[1 + \Delta]$, in dotted line transfer function of discretized ground model

Unstructured uncertainty is represented by the multiplicative dynamic perturbation Δ, known only to belong to the *modeling error set* Δ_E defined as

$$\Delta_E = \{\Delta : \|W^{-1}(\omega)\,\Delta(\omega)\|_\infty \le 1,\ 0 \le \omega < \infty\}, \tag{22}$$

where $W(\omega)$ is given by (20).

4. ROBUST REGULATOR DESIGN

The feedback control configuration shown in Fig. 3 has been considered for output regulation of the plant.

The loop transfer function

$$L(s, p, \Delta) = A(s)G(s, p)[1 + \Delta]M(s)C(s), \tag{23}$$

is now considered. Symbol $L(s, p, \Delta)$ denotes that the loop transfer function is affected by both structured and unstructured perturbations.

Furthermore, the sensitivity function $S(s, p, \Delta)$ and its complement $T(s, p, \Delta)$ are considered:

$$S = (1 + L)^{-1}, \quad T = 1 - S = L(1 + L)^{-1}. \tag{24}$$

The robust regulation problem requires to design a relatively low order controller $C(s)$ such that the following specifications are met:

(1) Closed loop is robustly stable with respect to the allowed structured and unstructured uncertainties, i e. $\forall p \in \Pi$ and $\forall \Delta \in \Delta_F$.

(2) Closed loop control guarantees disturbance attenuation of 1:100 in r.m.s. values with respect to open loop operation. This is obtained by imposing:

$$|S(j\omega, p, \Delta)| \le U(\omega), \quad \forall p \in \Pi, \quad \forall \Delta \in \Delta_E, \quad 0.63 \le \omega \le 57 \text{ rad sec}^{-1}, \tag{25}$$

where $U(\omega)$ is the bounding function shown in Fig. 8, and $[0.63, 57]$ rad sec^{-1} is the frequency range where the spectral density function $S_d(\omega)$ of the disturbance has more than 95% of its power (see Fig. 9).

(3) Command amplitude is limited, in order to guarantee that the current in the electromagnetic actuators does not exceed 10 A r.m.s. This is obtained by imposing:

$$|H(j\omega, p)S(j\omega, p, \Delta)M(j\omega)C(j\omega)A(j\omega)| \le 6 \times 10^6 \text{ Am}^{-1}\text{ sec}^2, \tag{26}$$

$$\forall p \in \Pi, \quad \forall \Delta \in \Delta_E, \quad 0.63 \le \omega \le 57 \text{ rad sec}^{-1}.$$

A systematic approach to a design problem of such a complexity is of the possibilities offered by the present state of the robust control literature. Consequently a design approach has been used, based on iterative phases of "nominal" design and robustness analysis, which, in case of failure, gives indications for the successive design phases.

The design phase has been carried on, using the nominal parameters, in the following way: the robust stability requirement with respect to unmodelled dynamics and the given performance

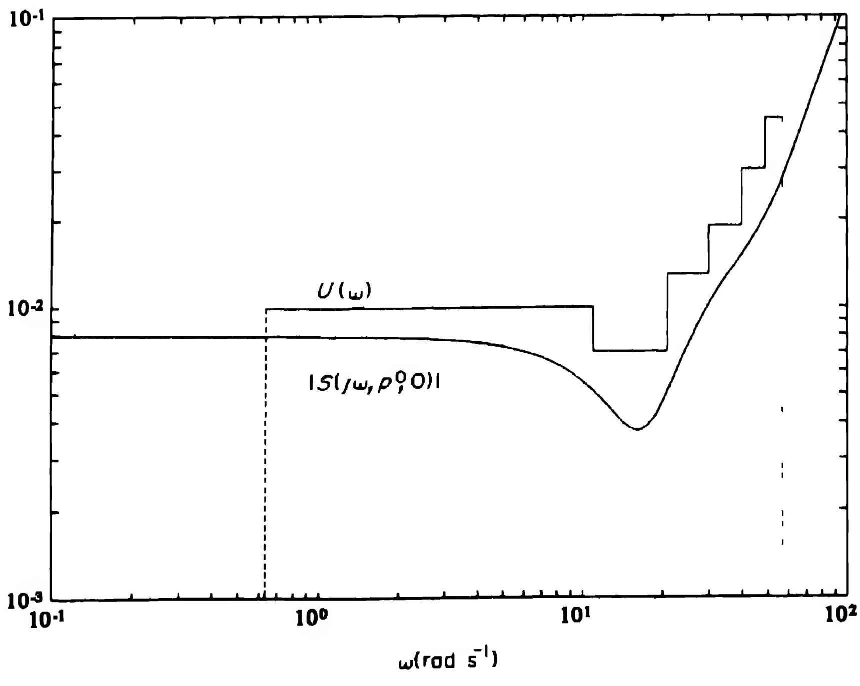


FIG. 8. Nominal sensitivity ($C_K(s) = C_{10}(s)$) and bounding functions.

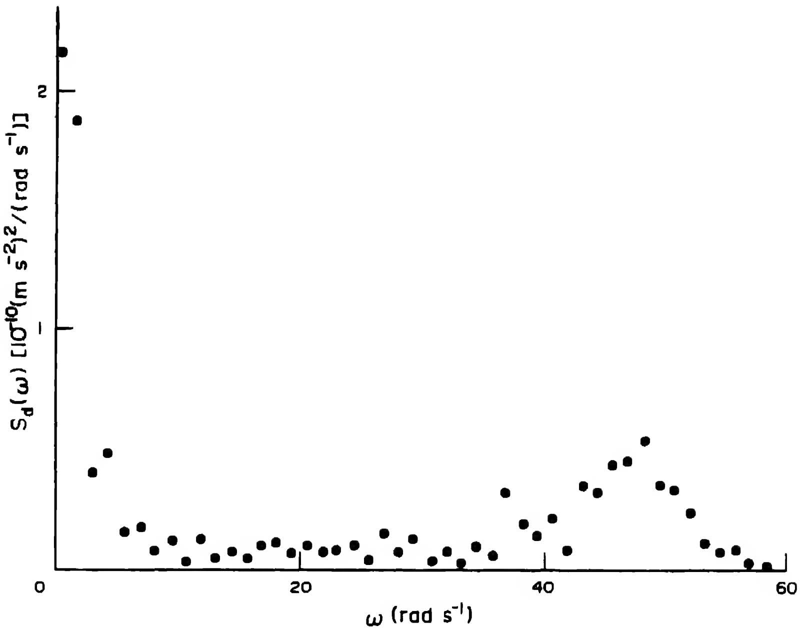


FIG. 9. Spectral density of disturbance.

specifications are used to suitably define the weighting functions entering in the H_∞ design approach. Through H_∞ synthesis algorithms it is possible to find a controller, satisfying the specifications for the "nominal" parameters, or to understand if some of the specifications have to be relaxed. Note that the controller obtained by H_∞ synthesis may be of higher order than desirable. However, it is often possible to find, by means of order reduction techniques, a simpler controller satisfying nevertheless the nominal performances. For this design phase we have used an interactive design software described in Ferro (1992), implemented using the Robust-Control Toolbox of MATLAB.

The robustness analysis phase has been carried on following the approach proposed by Milanese *et al.* (1991).

First, performance specifications are represented in the frequency domain by functions $F_k(\omega, p)$ $k = 1$–3 such that the kth performance specification is satisfied for all unstructured perturbations $\Delta \in \Delta_E$ if and only if $F_k(\omega, p) > 0$, $\omega \in \Omega_k$. For example, in the case of stability, the corresponding specification inequality can be obtained by the small gain theorem, giving:

$$|W(\omega)T(j\omega, p, 0)| < 1. \tag{27}$$

A given compensator is said to achieve robust k-performance if $F_k(\omega, p) > 0$, $\omega \in \Omega_k$, $\forall p \in \Pi$. Note that in this way, robustness is guaranteed vs both parametric and unstructured uncertainty.

A robustness measure for the kth performance, called performance margin ρ_k^*, is defined as the radius, according to l_∞^u norm in parameter space, of the maximal ball, centered at the nominal parameter p^0, such that the closed loop system preserved the given performance specification for every parameter vector belonging to the maximal ball and for all admissible unstructured perturbations. This measure is a generalization of the widely used concept of stability margin (see e.g. de Gaston and Safonov, 1988; Šiljak, 1989; Vicino *et al.*, 1990).

Performance margin ρ_k^* can be computed as

$$\rho_k{}^* = \min_{p,\omega,\rho} \rho$$
$$\text{subject to}$$
$$\begin{cases} \rho \geq 0 \\ |p_i - p_i^0| \leq \rho u_i \quad i = 1, \ldots, 4 \\ F_k(\omega, p) \leq 0, \quad \omega \in \Omega_k. \end{cases} \tag{28}$$

Note that the kth performance is robustly achieved if and only if $\rho_k{}^* \geq 1$.

The optimization problem (28) may have local minima and the global solution is needed for solving the problem. Global optimization methods based on random search algorithms are not appropriate, since these methods guarantee convergence to the global minimum only in probability and, more importantly, they do not give any measure on how far the obtained solution is from the true global minimum. When $F(\omega, p)$ is a polynomial function in ω and p, algorithms able to produce a sequence of upper and lower bounds converging with certainty to global extrema have been proposed, and shown

to be able to solve some non trivial robustness problems (Vicino *et al.*, 1990; Vicino and Milanese, 1990; Milanese *et al.*, 1991).

Indeed the given specifications can be represented by polynomial inequalities in p and ω. The explicit functional expressions $F_k(\omega, p)$, $k = 1$–3, corresponding to the considered performances, have been obtained by means of symbolic manipulation package DERIVE on a personal computer. As an example, $F_1(\omega, p)$ is reported in the Appendix.

The results reported in this paper have been obtained by use of the algorithm reported in Malan *et al.* (1992), being a further improvements over the previously cited ones.

Note that in case the controller is found not robustly performing (i.e. $\rho_k^* \leq 1$ for some k), the above analysis gives useful indications for the design phase. In particular, if the robustness margin of one performance is less than one, a new controller may be designed by strengthening the corresponding specification and possibly relaxing the specifications with robustness margin greater than one.

Following the described approach, a compensator has been found, satisfying performance specifications for nominal parameter p^0, given by:

$$\bar{C}(s) = 5\frac{(1+s/0.5)(1+s/11)(1+s/22)^2 \times (1+s/67)(1+s/405)(1+s/10^3)}{s^3(1+s/57)^2(1+s/58.8) \times (1+s/1065)(1+s/9\times 10^{10})}. \tag{29}$$

This transfer function has been found to be reasonably well approximated by the fourth order transfer function:

$$C_{10}(s) = 10\frac{(1+s/0.5)(1+s/50)^2(1+s/1000)}{s^3(1+s/200)} \tag{30}$$

This controller largely satisfies the given specifications in correspondence to the nominal model: the closed loop is stable with a damping factor of 0.4, the disturbance effect on the output is 1:120 with respect to the open loop, the current in the electromagnetic actuators does not exceed 7 A. However, as it results from Table 1, it does not achieve robust disturbance attenuation. This suggests that the gain loop has to be raised.

Indeed compensator $C_{14}(s) = 1.4C_{10}(s)$ has been found to achieve all robust performances. Note that the stability margin for this compensator is near to 1, indicating that higher gain may cause stability problems, as confirmed for example by the robustness analysis of $C_{20}(s) = 2C_{10}(s)$.

TABLE 1. SPECIFICATION MARGINS AND CORRESPONDING COMPUTING TIMES ON A VAX 9000 COMPUTER

Specification	$C_{10}(s)$ ρ^*	$C_{10}(s)$ CPU time (sec)	$C_{14}(s)$ ρ^*	$C_{14}(s)$ CPU time (sec)	$C_{20}(s)$ ρ^*	$C_{20}(s)$ CPU time (sec)
1	1.18	128	1.08	80	0.81	104
2	0.63	15	1.46	23	2.16	89
3	3.33	5	3.33	5	3.33	4

We remark that the robustness analysis has been carried on by an approach able to exploit the nonlinear structure induced by the considered perturbed parameters. It is known that this is an hard problem, requiring computationally cumbersome algorithms. Computationally simpler techniques have also been tried, by considering an interval plant formulation (Chappellat and Bhattacharyya, 1989; Tesi and Vicino, 1991). However this approach lead to conservative results, and for the present case study the resulting conservativeness is so high to prevent the possibility of finding a controller which could be guaranteed to perform robustly by such a simplified analysis (see Milanese *et al.*, 1992). On the other hand, for this case study, the computational burden of the nonlinear analysis appeared to be quite acceptable (see computing times in Table 1) and worth paying, in consideration of the improvements in the obtained results.

5 CONCLUSIONS

The presented case study illustrates how some robust identification and control techniques can be applied in dealing with real world problems.

The modeling and identification step has been performed by using physical insight and set membership identification theory, able to account for parametric variations and modeling approximations.

Robustness of the control system with respect to both types of uncertainty is considered, not for stability only, but for other performance specifications, such as disturbance attenuation and command power limitation.

An iterative design strategy has been adopted, based on successive phases of "nominal" design and robustness analysis.

The main conclusion that we draw from this case study is that, in facing with real world problems, it is necessary to take approaches with the following features: ability to deal with nonlinear physical parametrizations; accounting for both parametric uncertainty and unmodeled dynamics; designing and analysing with respect to different performances.

These requirements clearly lead to difficult problems, but it appears that techniques now exist, able to solve cases of such a complexity to be of some interest in practical applications

Acknowledgements—This work was partially supported by funds of Ministero dell'Università e della Ricerca Scientifica e Tecnologica and of CNR Special Project "Algorithms and Architectures for Identification and Robust and Adaptive Control".

REFERENCES

Belforte, G. and M. Milanese (1981). Uncertainty intervals evalution in presence of unknown but bounded errors. Nonlinear families of models. *Proc. 1st IASTED Symp. Modelling, Identification and Control*, Davos.

Chapellat, H. and S. P. Bhattacharyya (1989). A generalization of Kharitonov's theorem: robust stability of interval plants. *IEEE Trans. Aut. Control*, **34**, 306–311.

Dorato, P. and R. K. Yedavally (Eds) (1990). *Recent Advances in Robust Control.* IEEE Press.

Fan M. K. H., A. L. Tits and J. C. Doyle (1991). Robustness in presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Trans. Aut. Control*, **AC-36**, 25–38.

Ferro, M. (1992). Progetto assistito da calcolatore di un sistema di controllo robusto H_∞ sulla base di specifiche classiche. Relatori: G. Fiorio, S. Malan. *Tesi di Laurea*, Politecnico di Torino.

Fiorio, G., S. Malan, M. Milanese and A. Vicino (1990). Guaranteed specification control design of inertial platforms for high accuracy calibration device. *Proc. 29th IEEE Conf. on Decision and Control.*

de Gaston, R. R. E. and M. G. Safonov (1988). Exact calculation of the multiloop stability margin. *IEEE Trans. Aut. Control*, **AC-33**, 156–171.

Malan, M., M. Milanese, M. Taragna and J. Garloff (1992). B^3 Algorithm for robust performances analysis in presence of mixed parametric and dynamic perturbations. *Proc. 31st IEEE Conf. on Decision and Control.*

Milanese, M. and G. Belforte (1982). Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors. Linear families of models and estimators. *IEEE Trans. Aut. Control*, **AC-27**, 408–414.

Milanese, M. and R. Tempo (1985). Optimal algorithms theory for estimation and prediction. *IEEE Trans. Aut. Control*, **AC-30**, 730–738.

Milanese, M., R. Tempo and A. Vicino (Eds) (1989). *Robustness In Identification and Control.* Plenum Press, New York.

Milanese, M. and A. Vicino. (1991a). Estimation theory for nonlinear models and set membership uncertainty. *Automatica*, **25**, 403–408.

Milanese, M. and A. Vicino (1991b). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, **27**, 997–1009.

Milanese, M., G. Fiorio, S. Malan and A. Vicino (1991). Robust performance design of nonlinearly perturbed control systems. In S. P. Bhattacharya and L. Keel (Eds), *Robust Control.* CRC Press, Boca Raton, FL.

Milanese, M., M. Taragna, A. Trisoglio and S. Malan (1992). Linear vs nonlinear approaches to robustness analysis of LTI control systems. *Proc. Int. Workshop on Robust Control*, Ascona.

Schweppe, F. C. (1973). *Uncertain Dynamic Systems.* Prentice Hall, Englewood Cliffs, NJ.

Siljak, D. D. (1989). Parameter space methods for robust control design a guided tour. *IEEE Trans. Aut. Control*, **AC-34**, 674–688.

Tempo, R. and G. Wasilkowski (1988). Maximum likelihood estimators and worst case optimal algorithms for system identification. *Systems and Control Letters*, **10**, 265–270.

Tesi, A. and A. Vicino (1991). Kharitonov segments suffice for frequency response analysis of interval plant-controller. In *Proc. Int. Workshop on Robust Control*, San Antonio pp. 403–416 (U.S.A.). CRC Press, Boca Raton, FL.

Vicino, A. and M. Milanese (1990). Robust stability of linear state space models via Bernstein polynomials. In D. Hinrichsen and B. Martensson (Eds), *Control of Uncertain Systems.* Birkhauser, Basel.

Vicino, A., A. Tesi and M. Milanese (1990) Computation of non conservative stability perturbation bounds for systems with nonlinearly correlated uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 845–841.

APPENDIX

In order to show the type of functional expressions of functions $F_k(\omega, p)$, we report the specification function $F_1(\omega, p)$, related to robust stability of the closed loop system using controller $C_{10}(S)$.

As can be seen, $F_1(\omega, p)$ is composed of 101 monomials in variables ω, k_1, k_2, β_1, β_2, m_1. The other specification functions have approximately the same complexity. In particular $F_2(\omega, p)$ is composed of 71 monomials and $F_3(\omega, p)$ of 83 monomials.

We recall (see Milanese *et al.*, 1991) that the closed loop system is robustly stable $\forall p \in \Pi$, $\forall \Delta \in \Delta_E$ if and only if $F_1(\omega, p) > 0$, $\forall p \in \Pi$, $\forall \omega$.

$$
\begin{aligned}
F_1(\omega, p) = {} & \beta_1^2\beta_2^2\omega^8 + 26/25 \cdot 10^6\beta_1^2\beta_2^2\omega^6 + 4 \cdot 10^{10}\beta_1^2\beta_2^2\omega^4 + \\
& \beta_1^2k_2^2\omega^6 + 26/25 \cdot 10^6\beta_1^2k_2^2\omega^4 + 4 \cdot 10^{10}\beta_1^2k_2^2\omega^2 - \\
& 880\beta_1^2k_2\omega^8 - 1144/125 \cdot 10^8\beta_1^2k_2\omega^6 - 88/25 \cdot \\
& 10^{13}\beta_1^2k_2\omega^4 + 242/125 \cdot 10^5\beta_1^2\omega^{10} + 6292/3125 \cdot \\
& 10^{11}\beta_1^2\omega^8 + 968/125 \cdot 10^{15}\beta_1^2\omega^6 + 1566/625 \cdot \\
& 10^7\beta_1\beta_2^2\omega^8 + 39{,}933/15625 \cdot 10^{13}\beta_1\beta_2^2\omega^6 + \\
& 3132/625 \cdot 10^{17}\beta_1\beta_2^2\omega^4 + 3949/3125 \cdot 10^7\beta_1\beta_2\omega^{10} + \\
& 440{,}363/312500 \cdot 10^{13}\beta_1\beta_2\omega^8 + 113{,}179/78{,}125 \cdot \\
& 10^{18}\beta_1\beta_2\omega^6 - 3828/625 \cdot 10^{21}\beta_1\beta_2\omega^4 + 1566/625 \cdot \\
& 10^7\beta_1k_2^2\omega^6 + 39{,}933/15{,}625 \cdot 10^{13}\beta_1k_2^2\omega^4 + \\
& 3132/625 \cdot 10^{17}\beta_1k_2^2\omega^2 - 17{,}226/15{,}625 \cdot \\
& 10^{10}\beta_1k_2\omega^8 - 439{,}263/390{,}625 \cdot 10^{16}\beta_1k_2\omega^6 - \\
& 34{,}452/15{,}625 \cdot 10^{20}\beta_1k_2\omega^4 + \beta_2^2k_1^2\omega^6 + 26/25 \cdot \\
& 10^6\beta_2^2k_1^2\omega^4 + 4 \cdot 10^{10}\beta_2^2k_1^2\omega^2 - 2\beta_2^2k_1m_1\omega^8 - 52/25 \\
& \cdot 10^6\beta_2^2k_1m_1\omega^6 - 8 \cdot 10^{10}\beta_2^2k_1m_1\omega^4 - \\
& 28{,}720\beta_2^2k_1\omega^8 - 40{,}033/12{,}500 \cdot 10^{10}\beta_2^2k_1\omega^6 - \\
& 10{,}289/3125 \cdot 10^{15}\beta_2^2k_1\omega^4 + 174/125 \cdot 10^{19}\beta_2^2k_1\omega^2 \\
& + \beta_2^2m_1^2\omega^{10} + 26/25 \cdot 10^6\beta_2^2m_1^2\omega^8 + 4 \cdot 10^{10}\beta_2^2m_1^2\omega^6 \\
& + 28{,}720\beta_2^2m_1\omega^{10} + 40{,}033/12{,}500 \cdot 10^{10}\beta_2^2m_1\omega^8 + \\
& 10{,}289/3125 \cdot 10^{15}\beta_2^2m_1\omega^6 - 174/125 \cdot 10^{19}\beta_2^2m_1\omega^4 \\
& - 468{,}247/56{,}250 \cdot 10^7\beta_2^2\omega^{10} + \\
& 10{,}452{,}293/9{,}375{,}000 \cdot 10^{14}\beta_2^2\omega^8 + \\
& 30{,}577{,}679/15{,}625{,}000 \cdot 10^{20}\beta_2^2\omega^6 + \\
& 54{,}186{,}761/5{,}625{,}000 \cdot 10^{23}\beta_2^2\omega^4 + 7569/6250 \cdot \\
& 10^{27}\beta_2^2\omega^2 + 17{,}226/15{,}625 \cdot 10^{10}\beta_2k_1\omega^8 + \\
& 439{,}263/390{,}625 \cdot 10^{16}\beta_2k_1\omega^6 + 34{,}452/15{,}625 \cdot \\
& 10^{20}\beta_2k_1\omega^4 - 17{,}226/15{,}625 \cdot 10^{10}\beta_2m_1\omega^{10} - \\
& 439{,}263/390{,}625 \cdot 10^{16}\beta_2m_1\omega^8 - 34{,}452/15{,}625 \cdot \\
& 10^{20}\beta_2m_1\omega^6 + k_1^2k_2^2\omega^4 + 26/25 \cdot 10^6k_1^2k_2^2\omega^2 + 4 \cdot \\
& 10^{10}k_1^2k_2^2 - 880k_1^2k_2\omega^6 - 1144/125 \cdot 10^8k_1^2k_2\omega^4 - \\
& 88/25 \cdot 10^{13}k_1^2k_2\omega^2 + 242/125 \cdot 10^5k_1^2\omega^8 + \\
& 6292/3125 \cdot 10^{11}k_1^2\omega^6 + 968/125 \cdot 10^{15}k_1^2\omega^4 - \\
& 2k_1k_2^2m_1\omega^6 - 52/25 \cdot 10^6k_1k_2^2m_1\omega^4 - 8 \cdot \\
& 10^{10}k_1k_2^2m_1\omega^2 - 28{,}720k_1k_2^2\omega^6 - 40{,}033/12{,}500 \cdot \\
& 10^{10}k_1k_2^2\omega^4 - 10{,}289/3125 \cdot 10^{15}k_1k_2^2\omega^2 + 174/125 \\
& \cdot 10^{19}k_1k_2^2 + 1760k_1k_2m_1\omega^8 + 1144/625 \cdot \\
& 10^9k_1k_2m_1\omega^6 + 176/25 \cdot 10^{13}k_1k_2m_1\omega^4 + \\
& 3949/3125 \cdot 10^7k_1k_2\omega^8 + 440{,}363/312{,}500 \cdot \\
& 10^{13}k_1k_2\omega^6 + 113{,}179/125 \cdot 10^{18}k_1k_2\omega^4 - \\
& 3828/625 \cdot 10^{21}k_1k_2\omega^2 - 484/125 \cdot 10^5k_1m_1\omega^{10} - \\
& 12{,}584/3125 \cdot 10^{11}k_1m_1\omega^8 - 968/625 \cdot 10^{16}k_1m_1\omega^6 \\
& + k_2^2m_1^2\omega^8 + 26/25 \cdot 10^6k_2^2m_1^2\omega^6 + 4 \cdot 10^{10}k_2^2m_1^2\omega^4 + \\
& 28{,}720k_2^2m_1\omega^8 + 40{,}033/12{,}500 \cdot 10^{10}k_2^2m_1\omega^6 + \\
& 10{,}289/3125 \cdot 10^{15}k_2^2m_1\omega^4 - 174/125 \cdot 10^{19}k_2^2m_1\omega^2 \\
& - 468{,}247/56{,}250 \cdot 10^7k_2^2\omega^8 + \\
& 10{,}452{,}293/9{,}375{,}000 \cdot 10^{14}k_2^2\omega^6 + \\
& 30{,}577{,}679/15{,}625{,}000 \cdot 10^{20}k_2^2\omega^4 + \\
& 54{,}186{,}761/5{,}625{,}000 \cdot 10^{23}k_2^2\omega^2 + 7569/6250 \cdot \\
& 10^{27}k_2^2 - 880k_2m_1^2\omega^{10} - 1144/125 \cdot 10^8k_2m_1^2\omega^8 - \\
& 88/25 \cdot 10^{13}k_2m_1^2\omega^6 - 3949/3125 \cdot 10^7k_2m_1\omega^{10} - \\
& 440{,}363/312{,}500 \cdot 10^{13}k_2m_1\omega^8 - 113{,}179/78{,}125 \cdot \\
& 10^{18}k_2m_1\omega^6 + 3828/625 \cdot 10^{21}k_2m_1\omega^4 + 242/125 \cdot \\
& 10^5m_1^2\omega^{12} + 6292/3125 \cdot 10^{11}m_1^2\omega^{10} + 968/125 \cdot \\
& 10^{15}m_1^2\omega^8.
\end{aligned} \tag{31}
$$

Automatica, Vol. 29, No. 1, pp. 157–168, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

On the Design of Robust Two Degree of Freedom Controllers*

D. J. N. LIMEBEER,† E. M. KASENALLY† and J. D. PERKINS†

A new design method for two degree of freedom controllers is presented. $\mathcal{H}^\infty$ optimization methods are used to ensure a guaranteed level of robust stability and robust performance. The techniques are illustrated on a high purity distillation system.

Key Words—$\mathcal{H}^\infty$ control; two degrees of freedom; normalized coprime factors; loop shaping; robust performance; robust stability; model matching; control system design.

Abstract—The aim of this paper is to introduce two methods of designing robust two degree of freedom (TDF) controllers. Both design procedures will be illustrated and evaluated on a high purity distillation column example. In the first approach the feedback controller and the prefilter are designed in a single step using an $\mathcal{H}^\infty$ optimization procedure. The second method optimizes the feedback controller and prefilter in two separate design stages. Roughly speaking, the feedback controller is designed to meet robust stability and disturbance rejection specifications, while the prefilter is used to improve the robust model matching properties of the closed loop system. The single step approach has the advantages that it is easy to use, and that the resulting controller degree is the same as that of the plant. The two stage approach offers greater design flexibility and it may produce robust stability and robust performance margins which are significantly bigger than those achievable with the single stage approach. Set against that, the two stage technique is more difficult to use and it produces controllers which may have an order which is significantly bigger than that of the plant.

1. INTRODUCTION

In conventional control systems of the type illustrated in Fig. 1, the plant input u is generated by processing $e = r - \mathbf{F}y$ which is the difference between the reference inputs and a fixed function of the outputs. If the input to the controller is some fixed function of r and y (such as $e = r - y$ or $e = r - \mathbf{F}y$), it is well known that the associated design problem has a single degree of freedom (SDF) which may be described in terms of a stable $\mathbf{Q}$ parameter. In systems of this type it is also known that there are fundamental design compromises which have to be made between robust stability and sensitivity minimization (for example). When one is faced with demanding performance specifications, the restrictions associated with SDF control systems may make it difficult or even impossible to optimize the trade-offs which have to be made.

In two degree of freedom (TDF) configurations one allows the controller to process the references and measurements independently subject only to closed loop internal stability requirements. For example, one may set

$$u = \mathbf{K}\begin{bmatrix} r \\ y \end{bmatrix} = \mathbf{K}_1 r + \mathbf{K}_2 y, \qquad (1.1)$$

and then select $\mathbf{K}_1$ (the prefilter) and $\mathbf{K}_2$ (the feedback controller) arbitrarily with internal stability the only restriction. The two degrees of freedom in systems of this type may be parametrized in terms of two stable but otherwise free $\mathbf{Q}$ parameters (Vidyasagar, 1985; Youla and Bongiorno, 1985).

Linear quadratic Gaussian (LQG) methods and $\mathcal{H}^\infty$ optimization procedures developed over the last decade are now widely used in the design of SDF control systems. The aim or this paper is to investigate the possibility of expanding the role of $\mathcal{H}^\infty$ optimization tools in TDF system design. Our starting point is the TDF parametrization theory of Youla and Bongiorno (1985) and the loop shaping design procedure described in McFarlane and Glover (1992). We present two optimization procedures for the design of such systems. In the first approach the entire controller is designed in a single step. In this case we optimize the two $\mathbf{Q}$-parameters (which we will call $\mathbf{Q}_1$ and $\mathbf{Q}_2$) simultaneously. The second procedure settles on $\mathbf{Q}_2$ and therefore $\mathbf{K}_2$ first, and then optimizes $\mathbf{Q}_1$ and hence $\mathbf{K}_1$ in a second optimization step. The $\mathbf{Q}_2$

* Received 10 December 1991; received in final form 23 July 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor R. F. Curtain.

† Interdisciplinary Research Centre for Process Systems Engineering, Imperial College, Exhibition Road, London SW7 2BY, U.K.

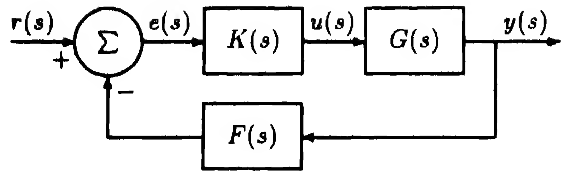


FIG. 1. The archetypal single degree of freedom controller configuration

parameter is used to optimize the loop's robust stability and disturbance rejection properties, while $\mathbf{Q}_1$ is used to ensure a prescribed level of robust model matching. The first procedure has the advantage that it is easy to use and that the resulting controller has a low degree because $\mathbf{K}_1$ and $\mathbf{K}_2$ share the same state-space. The second method will always have the capacity to produce a higher robust stability margin, and may result in an improved level of robust performance (as compared with the first approach). Unfortunately the two stage approach requires more computation than the first, is more difficult to use and results in a higher order controller. Both design procedures will be tested and compared on a high purity distillation system.

The theoretical background to the design methods is given in Section 2 where we summarize the notion of robust stability in the context of normalized coprime factors, robust stability optimization, a measure of robust model matching and TDF parametrization theory. The TDF design procedures are given in Section 3. A detailed design for the distillation problem appears in Section 4, with the conclusions in Section 5.

The space of functions with no poles in the closed right half plane is denoted $\mathcal{H}_\infty^+$. Bold face letters will be used for transfer functions and script characters will be used for sets or spaces. The notation

$$\mathbf{G} = D + C(sI - A)^{-1}B \stackrel{s}{=} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \quad (1.2)$$

will be used from time to time, and we also need $\mathbf{G}^{\sim}(s) := \mathbf{G}^*(-\bar{s})$. Linear fractional maps are denoted

$$\mathcal{F}_l(\mathbf{P}, \mathbf{K}) := \mathbf{P}_{11} + \mathbf{P}_{12}\mathbf{K}(\mathbf{I} - \mathbf{P}_{22}\mathbf{K})^{-1}\mathbf{P}_{21},$$

in which $\mathbf{P}$ is partitioned as $\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}$. Finally, we will make frequent use of one of Redheffer's theorems on the contractive properties of linear fractional maps:

Theorem 1.1. If $\|\mathbf{P}\|_\infty \le \gamma$ and $\|\mathbf{K}\|_\infty \le \gamma^{-1}$, then $\|\mathcal{F}_l(\mathbf{P}, \mathbf{K})\|_\infty \le \gamma$.

Proof. See Redheffer (1960). ■

2 THEORETICAL BACKGROUND

The problem of designing controllers to meet robust stability specifications together with nominal performance requirements has been well studied. One such method, which is based on normalized coprime factors, concentrates on coprime factor robustness optimization and has already been successfully applied to several design problems (McFarlane and Glover, 1992). The task of designing TDF controllers to simultaneously meet robust stability and robust performance specifications seems to have received less attention. In the remainder of this section we will summarize the theory required for two design procedures for TDF controllers. These procedures make it possible to guarantee a prescribed level of robust stability, while simultaneously ensuring some minimum level of robust model matching.

2.1. *Robust stability*

One way to characterize a set of plants, $\mathcal{G}$, with some nominal centre, $\mathbf{G}$, is via a perturbed normalized coprime factorization. If $\mathbf{G}$ is a given plant model, then

$$\mathbf{G} = \mathbf{M}^{-1}\mathbf{N}, \quad (2.1)$$

is a normalized left coprime factorization of $\mathbf{G}$ if $\mathbf{M}, \mathbf{N} \in \mathcal{H}_\infty^+$ are coprime and satisfy

$$\mathbf{M}\mathbf{M}^{\sim} + \mathbf{N}\mathbf{N}^{\sim} = \mathbf{I}. \quad (2.2)$$

If

$$\mathbf{G} \stackrel{s}{=} \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right],$$

is given

$$[\mathbf{N} \quad \mathbf{M}] \stackrel{s}{=} \left[\begin{array}{c|cc} A + HC & B + HD & H \\ \hline R^{-1/2}C & R^{-1/2}D & R^{-1/2} \end{array}\right], \quad (2.3)$$

is a normalized coprime factorization of $\mathbf{G}$ where $H := -(BD' + ZC')R^{-1}$ and $R := I + DD'$. The matrix $Z \ge 0$ is the unique stabilizing solution to the algebraic Riccati equation

$$(A - BS^{-1}D'C)Z + Z(A - BS^{-1}D'C)' - ZC'R^{-1}CZ + BS^{-1}B' = 0, \quad (2.4)$$

in which $S := I + D'D$.

A set of plants $\mathcal{G}$ may now be characterized by

$$\mathbf{G}_p = (\mathbf{M} - \Delta_M)^{-1}(\mathbf{N} + \Delta_N), \quad (2.5)$$

in which $\Delta_M, \Delta_N \in \mathcal{H}_\infty^+$ satisfy $\|\Delta_M \;\; \Delta_N\|_\infty < \gamma^{-1}$.

In order to maximize the robust stability of the closed loop system, we require a feedback

controller which minimizes

$$\gamma = \left\| \begin{bmatrix} \mathbf{K} \\ \mathbf{I} \end{bmatrix} (\mathbf{I} - \mathbf{GK})^{-1} \mathbf{M}^{-1} \right\|_\infty, \tag{2.6}$$

as **K** ranges over all stabilizing controllers. It is then immediate from the small gain theorem that the perturbed closed loop system will remain stable provided

$$\|\mathbf{\Delta}_M \quad \mathbf{\Delta}_N\|_\infty < \gamma^{-1}. \tag{2.7}$$

It is also well known that the lowest achievable $\mathscr{H}^\infty$ norm for the quantity in (2.6) is given by Glover and McFarlane (1989)

$$\gamma_{\text{opt}}^{-1} = \sqrt{1 - \|\mathbf{M} \quad \mathbf{N}\|_H^2}, \tag{2.8}$$

where $\|\cdot\|_H$ denotes the Hankel norm.

If $X \geq 0$ is the unique stabilizing solution to the Riccati equation

$$(A - BS^{-1}D'C)'X + X(A - BS^{-1}D'C) - XBS^{-1}B'X + C'R^{-1}C = 0, \tag{2.9}$$

it can be shown that

$$\gamma_{\text{opt}} = (1 + \lambda_{\max}(XZ))^{1/2}. \tag{2.10}$$

A controller $\mathbf{K}_0$ which achieves this bound is described by the generalized state-space equations (Glover and McFarlane, 1989)

$$Q'\dot{x} = [Q'(A + BF) + \gamma^2 ZC'(C + DF)]x + \gamma^2 ZC'u, \tag{2.11}$$

$$y = B'Xx - D'u, \tag{2.12}$$

where

$$F = -S^{-1}(D'C + B'X), \tag{2.13}$$

$$Q = (1 - \gamma^2)I + XZ. \tag{2.14}$$

In practical design applications it is usually necessary to reshape the plant's frequency response in order to meet the closed loop performance requirements. The loop shaping is done by premultiplying the plant by a loop shaping precompensator **W** (McFarlane and Glover, 1992). The robustness maximizing controller is then found for **GW** and it is consequently **GW** and not **G** which is optimally robustly stabilized.

2.2. *Robust performance*

The goal of this section is to develop a method of guaranteeing closed loop robust performance in a model matching sense in the face of nominal plant model perturbations. We will analyse the case of multiplicative perturbations at the plant input in detail, normalized coprime factor perturbations, additive perturbations and multiplicative perturbations at the plant output may then be dealt with in the same way. In the context of this paper, we will call the system robust from a model matching point of view if

$$\|(\mathbf{I} - \mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1 - \mathbf{M}_0\|_\infty \leq \gamma, \tag{2.15}$$

for all perturbed plant models $\mathbf{G}_p \in \mathscr{G}$. The transfer function $\mathbf{M}_0$ represents some desired closed loop transfer function and is chosen by the designer. In our applications, we will use $\mathbf{M}_0$ to explicitly introduce time domain specifications into the design process. In the case of multiplicative perturbations at the plant input, the set $\mathscr{G}$ is generated by $\mathbf{G}_p = \mathbf{G}(\mathbf{I} + \mathbf{\Delta})$ with $\mathbf{\Delta}$ chosen so that **G** and $\mathbf{G}_p$ have the same number of right half plane poles and such that $\|\mathbf{\Delta}\|_\infty < \gamma^{-1}$. Direct calculation gives:

$$\begin{aligned}
&(\mathbf{I} - \mathbf{G}(\mathbf{I} + \Delta)\mathbf{K}_2)^{-1}\mathbf{G}(\mathbf{I} + \Delta)\mathbf{K}_1 - \mathbf{M}_0 \\
&\quad = (\mathbf{I} - (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{G}\Delta\mathbf{K}_2)^{-1} \\
&\qquad \times (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{G}(\mathbf{I} + \Delta)\mathbf{K}_1 - \mathbf{M}_0 \\
&\quad = (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{GK}_1 - \mathbf{M}_0 \\
&\qquad + (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{G}\Delta \\
&\qquad \times (\mathbf{I} - (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_2\mathbf{G}\Delta)^{-1} \\
&\qquad \times (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 \\
&\quad = \mathscr{F}_l\left(\begin{bmatrix} (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{GK}_1 - \mathbf{M}_0 & (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{G} \\ (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_2\mathbf{G} \end{bmatrix}, \Delta\right).
\end{aligned}$$

If $\mathbf{K}_1$ and $\mathbf{K}_2$ can be found such that

$$\left\| \begin{matrix} (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{GK}_1 - \mathbf{M}_0 & (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{G} \\ (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_2\mathbf{G} \end{matrix} \right\|_\infty \leq \gamma, \tag{2.16}$$

then it is clear that:

(1) The feedback system will be stable for all those perturbations Δ which give **G** and $\mathbf{G}_p$ the same number of right half plane poles and such that $\|\Delta\|_\infty < \gamma^{-1}$. This follows from $\|(\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_2\mathbf{G}\|_\infty \leq \gamma$ and the small gain theorem.

(2) The closed loop system will have robust model matching properties in the sense that

$$\|(\mathbf{I} - \mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1 - \mathbf{M}_0\|_\infty \leq \gamma, \tag{2.17}$$

for all $\mathbf{G}_p \in \mathscr{G}$. This follows from $\|\Delta\|_\infty < \gamma^{-1}$, (2.16) and Redheffer's theorem on the contractive properties of linear fractional maps.

We will now repeat this analysis for perturbed normalized coprime factors. In this case each element $\mathbf{G}_p \in \mathscr{G}$ is represented by

$$\mathbf{G}_p = (\mathbf{M} - \mathbf{\Delta}_\mathbf{M})^{-1}(\mathbf{N} + \mathbf{\Delta}_\mathbf{N}), \tag{2.18}$$

where $\mathbf{\Delta}_\mathbf{M}, \mathbf{\Delta}_\mathbf{N} \in \mathscr{H}_\infty^+$ satisfy $\|\mathbf{\Delta}_\mathbf{M} \quad \mathbf{\Delta}_\mathbf{N}\|_\infty < \gamma^{-1}$.

Suppose

$$\Phi = (\mathbf{I} - (\mathbf{M} - \Delta_M)^{-1}(\mathbf{N} + \Delta_N)\mathbf{K}_2)^{-1} \times (\mathbf{M} - \Delta_M)^{-1}(\mathbf{N} + \Delta_N)\mathbf{K}_1 - \mathbf{M}_0, \quad (2.19)$$

in which $\mathbf{K}_1$ and $\mathbf{K}_2$ are given, and where Δ_M, $\Delta_N \in \mathcal{H}_\infty^+$ satisfy $\|\Delta_M \;\; \Delta_N\|_\infty < \gamma^{-1}$. Then a rearrangement of the type given earlier yields

$$\Phi = \mathcal{F}_l\left(\left[\begin{array}{c|c} (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1 - \mathbf{M}_0 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \hline \mathbf{G}(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ (\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & \mathbf{K}_2(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \end{array}\right], [\Delta_M \;\; \Delta_N]\right). \quad (2.20)$$

From this we observe that

$$\left\|\left[\begin{array}{c|c} (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1 - \mathbf{M}_0 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \hline \mathbf{G}(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ (\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & \mathbf{K}_2(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \end{array}\right]\right\|_\infty \le \gamma, \quad (2.21)$$

and

$$\|\Delta_M \;\; \Delta_N\|_\infty < \gamma^{-1}, \quad (2.22)$$

ensure that:

(1) The closed loop system in Fig. 2 will be stable for every $\mathbf{G}_p \in \mathcal{G}$. This follows from the fact that the (2, 2) block of (2.21) is stable, has infinity norm $\le\gamma$ and the small gain theorem.
(2) The closed loop has the robust model matching property

$$\|(\mathbf{I} - \mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1 - \mathbf{M}_0\|_\infty \le \gamma, \quad (2.23)$$

for all $\mathbf{G}_p \in \mathcal{G}$. This property is immediate from Redheffer's theorem (Redheffer 1960).

2.3. *All stabilizing controllers*

As we will now demonstrate, the TDF controller may be parametrized in terms of a pair of Q-parameters which we will denote $\mathbf{Q}_1$ and $\mathbf{Q}_2$. A theory of this type is well known, and may be found in Vidyasagar (1985) and Youla and Bongiorno (1985). Once all the controllers have been parametrized, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ may be optimized either separately or in a single step. We will study both these approaches in the context of the TDF configuration in Fig. 2.

It follows from this diagram that

$$\begin{bmatrix} z \\ y \\ u \\ \hline r \\ y \end{bmatrix} = \left[\begin{array}{cc|c} -\mathbf{M}_0 & \mathbf{M}^{-1} & \mathbf{G} \\ \mathbf{0} & \mathbf{M}^{-1} & \mathbf{G} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \hline \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{-1} & \mathbf{G} \end{array}\right]\begin{bmatrix} r \\ \phi \\ u \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}\begin{bmatrix} r \\ \phi \\ \hline u \end{bmatrix}. \quad (2.24)$$

We now suppose that

$$\mathbf{G} = \mathbf{M}^{-1}\mathbf{N} = \tilde{\mathbf{N}}\tilde{\mathbf{M}}^{-1}, \quad (2.25)$$

are left and right coprime factorizations of $\mathbf{G}$, that

$$\begin{bmatrix} \tilde{\mathbf{V}} & \tilde{\mathbf{U}} \\ -\mathbf{N} & \mathbf{M} \end{bmatrix}\begin{bmatrix} \tilde{\mathbf{M}} & -\mathbf{U} \\ \tilde{\mathbf{N}} & \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (2.26)$$

are the corresponding Bezout identities and that

$$\hat{\mathbf{K}} = -(\mathbf{U} + \tilde{\mathbf{M}}\hat{\mathbf{Q}})(\mathbf{V} - \tilde{\mathbf{N}}\hat{\mathbf{Q}})^{-1}, \quad \hat{\mathbf{Q}} \in \mathcal{H}_\infty^+, \quad (2.27)$$

is a representation formula for every feedback controller which stabilizes $\mathbf{G}$ (Glover *et al.*, 1991; Youla *et al.*, 1976). In the case that $\mathbf{M}_0$ is stable, $\mathbf{K} = [\mathbf{K}_1 \;\; \mathbf{K}_2]$ will stabilize $\mathbf{P}_{22}$ if and only if it stabilizes $\mathbf{P}$. Now

$$\mathbf{P}_{22} = \begin{bmatrix} \mathbf{0} \\ \mathbf{G} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{0} \\ \mathbf{N} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{N}} \end{bmatrix}\tilde{\mathbf{M}}^{-1}, \quad (2.28)$$

while the corresponding Bezout identities are

$$\left[\begin{array}{c|c|c} \tilde{\mathbf{V}} & \mathbf{0} & \tilde{\mathbf{U}} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} \\ -\mathbf{N} & \mathbf{0} & \mathbf{M} \end{array}\right]\left[\begin{array}{c|c|c} \tilde{\mathbf{M}} & \mathbf{0} & -\mathbf{U} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \tilde{\mathbf{N}} & \mathbf{0} & \mathbf{V} \end{array}\right] = \left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{array}\right]. \quad (2.29)$$

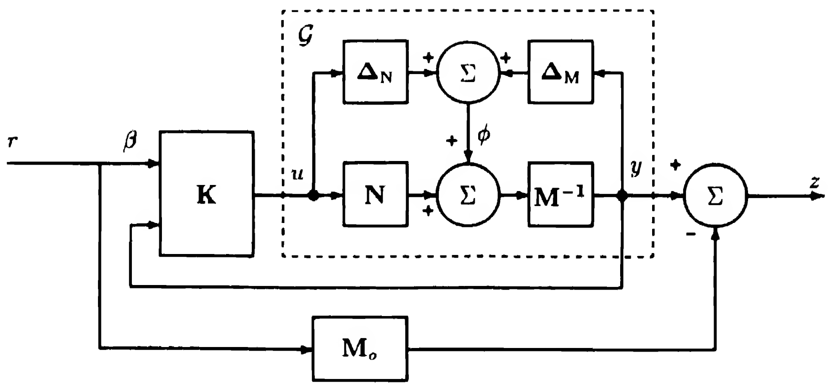


FIG. 2. TDF controller and a perturbed normalized coprime factor representation of $\mathcal{G}$.

As a consequence, we have that

$$[\mathbf{K}_1 \quad \mathbf{K}_2] = -[\tilde{\mathbf{M}}\mathbf{Q}_1 \quad \mathbf{U}+\tilde{\mathbf{M}}\mathbf{Q}_2] \times \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{N}} \end{bmatrix} [\mathbf{Q}_1 \quad \mathbf{Q}_2] \right\}^{-1}$$

$$= -[\tilde{\mathbf{M}}\mathbf{Q}_1 \quad \mathbf{U}+\tilde{\mathbf{M}}\mathbf{Q}_2] \times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ (\mathbf{V}-\tilde{\mathbf{N}}\mathbf{Q}_2)^{-1}\tilde{\mathbf{N}}\mathbf{Q}_1 & (\mathbf{V}-\tilde{\mathbf{N}}\mathbf{Q}_2)^{-1} \end{bmatrix}.$$

From this we conclude that

$$\mathbf{K}_2 = -(\mathbf{U}+\tilde{\mathbf{M}}\mathbf{Q}_2)(\mathbf{V}-\tilde{\mathbf{N}}\mathbf{Q}_2)^{-1}, \quad (2.30)$$

and

$$\mathbf{K}_1 = -(\tilde{\mathbf{M}}-\mathbf{K}_2\tilde{\mathbf{N}})\mathbf{Q}_1. \quad (2.31)$$

Substituting (2.30) and (2.31) into (2.21) gives

$$\left[\begin{array}{c|c} (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1-\mathbf{M}_0 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \hline \mathbf{G}(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ (\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & \mathbf{K}_2(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \end{array}\right] = \left[\begin{array}{c|c} -\mathbf{M}_0 & \mathbf{V} \\ \hline \mathbf{0} & \mathbf{V} \\ \mathbf{0} & -\mathbf{U} \end{array}\right] - \left[\begin{array}{c} \tilde{\mathbf{N}} \\ \hline \tilde{\mathbf{N}} \\ \tilde{\mathbf{M}} \end{array}\right] [\mathbf{Q}_1 \,|\, \mathbf{Q}_2], \quad (2.32)$$

which is affine in $\mathbf{Q}_1$ and $\mathbf{Q}_2$.

3. DESIGN PROCEDURES FOR TDF CONTROLLERS

We propose two design methods which will be tested on a high purity distillation system. The first technique uses a single step optimization procedure, while the second approach optimizes the parameters $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in separate steps.

3.1. *The single step approach*

The single step approach makes use of the configuration illustrated in Fig. 3. By updating (2.32) to represent Fig. 3, we obtain

$$\begin{bmatrix} z \\ y \\ u \end{bmatrix} = \left[\begin{array}{c|c} \rho^2((\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1-\mathbf{M}_0) & \rho(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \hline \rho\mathbf{G}(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \rho(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & \mathbf{K}_2(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \end{array}\right] \times \begin{bmatrix} r \\ \phi \end{bmatrix}, \quad (3.33)$$

in which the scaling factor ρ is used to weight the relative importance of robust stability as compared with robust model matching. If

$$\left\| \left[\begin{array}{c|c} \rho^2((\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1-\mathbf{M}_0) & \rho(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \hline \rho\mathbf{G}(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \\ \rho(\mathbf{I}-\mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 & \mathbf{K}_2(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{M}^{-1} \end{array}\right] \right\|_\infty \le \gamma, \quad (3.34)$$

then:

(1) The loop will remain stable for all Δ_M, $\Delta_N \in \mathcal{H}_\infty^+$ such that $\|\Delta_M \quad \Delta_N\|_\infty < \gamma^{-1}$. This follows from the (2, 2) partition of (3.33), (3.34) and the small gain theorem.
(2) A direct consequence of Redheffer's Theorem is $\|(\mathbf{I}-\mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1-\mathbf{M}_0\|_\infty \le \gamma\rho^{-2}$ for all $\mathbf{G}_p \in \mathcal{G}$ generated by $\Delta_M, \Delta_N \in \mathcal{H}_\infty^+$ such that $\|\Delta_M \quad \Delta_N\|_\infty < \gamma^{-1}$. This is the guaranteed robust performance property.
(3) If ρ is set to zero, the TDF problem reduces to the ordinary robust stability problem described earlier in Section 2.1.

There is an alternative scaling procedure which may prove more useful in certain applications. We can perform all the "ρ scaling" through $\mathbf{M}_0$ by setting the (1, 1) partition of (3.33) to $((\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1-\rho^2\mathbf{M}_0)$. Once the optimization problem has been solved, we reverse the scaling by replacing $\mathbf{K}_1$ with $\rho^{-2}\mathbf{K}_1$; increasing ρ tends to emphasise the model matching part of

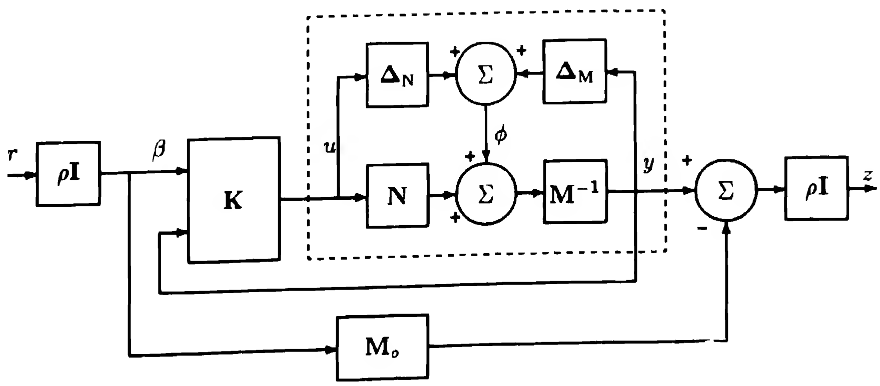


FIG. 3. The single step TDF design configuration.

the problem. In this case it follows that $\|(\mathbf{I}-\mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1-\rho^2\mathbf{M}_0\|_\infty \le \gamma$ for all $\mathbf{G}_p \in \mathcal{G}$.

As a refinement, the prefilter may be scaled so that the closed loop transfer function $\mathbf{R}_{y\beta}$ which maps $\beta \to y$ in Fig. 3 matches the model exactly at steady state. To do this we make the substitution $\mathbf{K}_1 \to \mathbf{K}_1 S$ where S is a scaling matrix defined by $S := \mathbf{R}_{y\beta}(0)^{-1}\mathbf{M}_0(0)$. We have observed that this rescaling has the effect of producing better model matching at all frequencies because the $\mathcal{H}_\infty$ optimization process tends to give $\mathbf{R}_{y\beta}$ the same magnitude frequency response shape as the model $\mathbf{M}_0$.

To set the problem up in a generalized regulator framework for $\mathcal{H}_\infty$ optimization (Glover *et al.*, 1991), we define the matrix

$$\left[\begin{array}{c} z \\ y \\ u \\ \hline \beta \\ y \end{array}\right] = \left[\begin{array}{cc|c} -\rho^2\mathbf{M}_0 & \rho\mathbf{M}^{-1} & \rho\mathbf{G} \\ \mathbf{0} & \mathbf{M}^{-1} & \mathbf{G} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \hline \rho\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{-1} & \mathbf{G} \end{array}\right] \tag{3.35}$$

which comes from Fig. 3. Setting

$$\mathbf{M}_0 \stackrel{s}{=} \left[\begin{array}{c|c} A_0 & B_0 \\ \hline C_0 & D_0 \end{array}\right] \quad \text{and}$$

with $\mathbf{M}_0$ chosen stable, gives

$$\left[\begin{array}{c|c} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \hline \mathbf{P}_{21} & \mathbf{P}_{22} \end{array}\right] \stackrel{s}{=} \left[\begin{array}{cc||cc|c} A & 0 & 0 & (BD'+ZC')R^{-1/2} & B \\ 0 & A_0 & -B_0 & 0 & 0 \\ \hline\hline \rho C & \rho^2 C_0 & -\rho^2 D_0 & \rho R^{1/2} & \rho D \\ C & 0 & 0 & R^{1/2} & D \\ 0 & 0 & 0 & 0 & I \\ \hline 0 & 0 & \rho I & 0 & 0 \\ C & 0 & 0 & R^{1/2} & D \end{array}\right] \tag{3.36}$$

which may be passed to standard software packages.

3.1.1. *A prescriptive design procedure.* As with any control system design methodology the process is iterative. As a starting point we assume that we have a nominal model of the plant together with some characterization of uncertainty. We then proceed as follows:

(1) Select a loop shaping weight for the open loop plant. This is used to meet certain closed loop performance specifications (McFarlane and Glover, 1992).
(2) Select simple target model, $\mathbf{M}_0$, for the closed loop system. This is usually a diagonal matrix of first or second order lags which represent desired closed loop time domain properties. The selected target model must be realistic, or the resulting closed loop system will have poor robust stability properties and the controller will produce excessive control signals.
(3) Find the minimal value γ_{opt} in the pure robust stabilization problem; this may be calculated using equation (2.10). A high value of γ_{opt} indicates that the specified loop shapes are inconsistent with robust stability requirements, and that $\mathbf{K}_0$ will significantly alter the loop shapes. In this case the loop shapes should be adjusted.
(4) Set ρ for the TDF problem in (3.35). For the high purity distillation problem we used $1 \le \rho \le 3$.
(5) Find the optimal value of γ. In the distillation problem application we found that $1.2\gamma_{opt} \le \gamma \le 3\gamma_{opt}$ gave a good compromise between the robust stability and robust performance objectives.
(6) Calculate the optimal controller, post multiply it by $\mathbf{W}$, and rescale the prefilter to achieve perfect steady model matching. The final controller degree will be $\le \deg(\mathbf{G}) + \deg(\mathbf{M}_0) + 2\deg(\mathbf{W})$.

3.2. *The two stage approach*

There are many ways of utilizing the two degrees of freedom offered in a TDF feedback configuration. The procedure we will now describe is aimed at guaranteeing robust stability in the face of coprime factor uncertainty and robust model matching in the presence of multiplicative perturbations at the plant input. This compromise gives us the advantages of the McFarlane–Glover loop shifting procedure for the design of $\mathbf{K}_2$, while explicitly taking account of the multiplicative form of the modelling uncertainty when designing the prefilter. In the case of the distillation column application, it is easy to show that the uncertain actuator gains and input time delays may be represented as stable uncertainty at the plant input. Direct calculation, using the modelling assumptions given in Section 4, shows that $\mathbf{G}_p = \mathbf{G}(\mathbf{I}+\mathbf{\Delta})$ where

$$\Delta = \begin{bmatrix} k_1 e^{-s(\tau-1)} - 1 & 0 \\ 0 & k_2 e^{-s(\tau-1)} - 1 \end{bmatrix}.$$

In the first step of the two step approach we seek a $\mathbf{K}_2$ which ensures some minimum level of robust stability. In particular we use $\mathbf{K}_2$ to ensure that

$$\left\| \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_2 \end{bmatrix} (\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1} [\mathbf{I} \quad \mathbf{G}] \right\|_\infty \le \gamma_2. \tag{3.37}$$

The second column of this expression appears as

the second column of (2.16) which is the linear fractional expression for robust model matching in the case of multiplicative perturbations at the plant input. Since

$$\begin{bmatrix} \tilde{\mathbf{M}} \\ \tilde{\mathbf{N}} \end{bmatrix} \stackrel{s}{=} \left[\begin{array}{c|c} A+BF & BS^{-1/2} \\ \hline F & S^{-1/2} \\ C+DF & DS^{-1/2} \end{array}\right], \tag{3.38}$$

is a normalized coprime factorization of $\mathbf{G}$, we may use the expression for $\mathbf{K}_2$ given in (2.11) and (2.12) together with $\mathbf{K}_1 = -(\tilde{\mathbf{M}} - \mathbf{K}_2\tilde{\mathbf{N}})\mathbf{Q}_1$ to parametrize all TDF controllers which robustly stabilize the system in Fig. 2. If

$$\mathbf{Q}_1 \stackrel{s}{=} \left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_2 \end{array}\right], \tag{3.39}$$

it follows by direct computation that $[\mathbf{K}_1 \quad \mathbf{K}_2]$ is described by the descriptor equations

$$\begin{aligned} Q'\dot{x}_1 &= (Q'(A+BF) + \gamma^2 ZC'(C+DF))x_1 \\ &\quad + (Q'B + \gamma^2 ZC'D)S^{-1/2}C_1x_2 \\ &\quad + (Q'B + \gamma^2 ZC'D)S^{-1/2}D_1u_1 + \gamma^2 ZC'u_2, \\ \dot{x}_2 &= A_1x_2 + B_1u_1, \\ y &= B'Xx_1 - S^{1/2}C_1x_2 - S^{1/2}D_1u_1 - D'u_2. \end{aligned}$$

The aim of the second step is to find a $\mathbf{Q}_1$ which enhances the robust command following performance. Since the uncertainty in the distillation column system appears at the plant input, we will seek a $\mathbf{Q}_1$ which minimizes

$$\gamma_1 = \left\| \begin{bmatrix} (\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{GK}_1 - \mathbf{M}_0 \\ (\mathbf{I} - \mathbf{K}_2\mathbf{G})^{-1}\mathbf{K}_1 \end{bmatrix} \right\|_\infty \tag{3.40}$$

$$= \left\| \begin{bmatrix} \mathbf{M}_0 + \tilde{\mathbf{N}}\mathbf{Q}_1 \\ \tilde{\mathbf{M}}\mathbf{Q}_1 \end{bmatrix} \right\|_\infty, \tag{3.41}$$

which is the quantity in the first column of (2.16). To summarize, $\mathbf{Q}_1 \in \mathcal{H}_\infty^+$ and $\mathbf{Q}_2 \in \mathcal{H}_\infty^+$ have been chosen to minimize the first and second columns, respectively of (2.16). Our aim is to select $\mathbf{Q}_1$ and $\mathbf{Q}_2$ so that

$$\left\| \begin{bmatrix} \mathbf{P}_{11} \\ \mathbf{P}_{21} \end{bmatrix} \right\|_\infty \le \gamma_1 \quad \left\| \begin{bmatrix} \mathbf{P}_{12} \\ \mathbf{P}_{22} \end{bmatrix} \right\|_\infty \le \gamma_2, \tag{3.42}$$

in which $\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}$ represents the expression in equation (2.16). It is now immediate that the loop will be robustly stable for all $\mathbf{\Delta}_\mathrm{N}, \mathbf{\Delta}_\mathrm{N} \in \mathcal{H}_\infty^+$ such that

$$\|\mathbf{\Delta}_\mathrm{M} \quad \mathbf{\Delta}_\mathrm{N}\|_\infty < \gamma_2^{-1}. \tag{3.43}$$

It is also clear that $\|(\mathbf{I} - \mathbf{GK}_2)^{-1}\mathbf{GK}_1 - \mathbf{M}_0\|_\infty \le \gamma_1$ which is a measure of the nominal performance. Our next result establishes the robust command following performance of the closed loop system in the sense that

$$\|(\mathbf{I} - \mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1 - \mathbf{M}_0\|_\infty \le \sqrt{\gamma_1^2 + \gamma_2^2}, \tag{3.44}$$

for all input multiplicative perturbations $\mathbf{\Delta} \in \mathcal{H}_\infty^+$ satisfying

$$\|\mathbf{\Delta}\|_\infty \le \frac{1}{\sqrt{\gamma_1^2 + \gamma_2^2}}. \tag{3.45}$$

Lemma 3.1. Suppose $\left\| \begin{bmatrix} \mathbf{P}_{11} \\ \mathbf{P}_{21} \end{bmatrix} \right\|_\infty \le \gamma_1$ and that $\left\| \begin{bmatrix} \mathbf{P}_{12} \\ \mathbf{P}_{22} \end{bmatrix} \right\|_\infty \le \gamma_2$. Then $\left\| \mathcal{F}_l\left(\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}, \mathbf{\Delta} \right) \right\|_\infty \le \sqrt{\gamma_1^2 + \gamma_2^2}$ for any $\mathbf{\Delta}$ such that $\|\mathbf{\Delta}\|_\infty \le \dfrac{1}{\sqrt{\gamma_1^2 + \gamma_2^2}}$.

Proof. To start, we note that $\|\mathbf{\Delta}\|_\infty \le \dfrac{1}{\sqrt{\gamma_1^2 + \gamma_2^2}}$ ensures that the closed loop is well posed for any $\gamma_1 > 0$. If

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix}, \tag{3.46}$$

we must have

$$\|z\|_2^2 + \|y\|_2^2 \le (\gamma_1^2 + \gamma_2^2)(\|u\|_2^2 + \|w\|_2^2), \tag{3.47}$$

since

$$\left\| \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \right\|_\infty \le \sqrt{\gamma_1^2 + \gamma_2^2}. \tag{3.48}$$

As $u = \mathbf{\Delta}y$ and $\|\mathbf{\Delta}\|_\infty \le \dfrac{1}{\sqrt{\gamma_1^2 + \gamma_2^2}}$, we conclude that $\|z\|_2 \le \sqrt{\gamma_1^2 + \gamma_2^2}\,\|w\|_2$ which completes the proof. ■

4 A HIGH PURITY DISTILLATION SYSTEM

In the remainder of this paper we compare and evaluate our TDF design methodologies on a high purity distillation system of the type described in Limebeer (1991), Morari and Zafiriou (1989), Skogestad *et al.* (1988) and Yaniv and Barlev (1990). Our study will be based on the following linearized model of the column

$$\mathbf{G}_p = \frac{1}{75s+1} \begin{bmatrix} 0.878 & -0.864 \\ 1.082 & -1.096 \end{bmatrix} \begin{bmatrix} k_1e^{-\tau_1 s} & 0 \\ 0 & k_2e^{-\tau_2 s} \end{bmatrix}. \tag{4.1}$$

The input time delays model uncertain flow dynamics in the column and lie in the range $0 \le \tau_1,\ \tau_2 \le 1$ minute. The actuator gains are also uncertain due to flow rate measurement errors. These variables lie in the range $0.8 \le k_1,\ k_2 \le 1.2$.

4.1. *Design specification*

The aim is to produce a controller which meets typical robust performance and robust stability specification. In our context robust

stability means guaranteed closed loop stability for all $0 \le \tau_1,\ \tau_2 \le 1$ minute and all $0.8 \le k_1,\ k_2 \le 1.2$. The robust performance specification is given in terms of a closed loop step response requirement to changes in the product compositions which must be met for all values of k_1, k_2, τ_1 and τ_2.

(1) For all product compositions where a change is demanded, the composition should be within ±10% of its desired value within 30 minutes, and then remain there for all future times.
(2) The final values of all variables should be within ±1% of their desired values.
(3) For product compositions where no change is demanded, deviations of less than ±50% at all times are required.

Two cases will be considered. In the first, a step demand $H(t)\begin{bmatrix}1\\0\end{bmatrix}$† is applied to the closed loop system. This input corresponds to a change in a relatively low gain direction of the plant (Skogestad *et al.*, 1988) and one would therefore expect some difficulty with control. In the second case we consider $H(t)\begin{bmatrix}0.4\\0.6\end{bmatrix}$ which is in a relatively high gain direction; no control difficulty is expected.

In conclusion, we mention that although we are concentrating on linear design, the final controller has to work on highly nonlinear hardware. As a consequence, the final controller should have a realistic gain and bandwidth.

4.2. *Distillation column design study*

We will tackle the distillation column design problem by applying the methods of Sections 3.1 and 3.2 to the nominal plant model

$$\mathbf{G} = \frac{e^{-s}}{75s+1}\begin{bmatrix}0.878 & -0.864\\ 1.082 & -1.096\end{bmatrix}, \qquad (4.2)$$

which corresponds to gain settings of $k_1 = k_2 = 1.0$ and $\tau_1 = \tau_2 = 1$ minute. The model used for the design is obtained by approximating the time delay by a first order Padé approximation. All our simulation studies use a sixth order Padé approximation for the delays. Once the controller has been found, it will be evaluated for its robust stability and robust performance characteristics. The remainder of this section deals with the single step approach, while Section 4.3 considers the two step approach for the same problem specification. Following the prescriptive design procedure given in Section 3.1.1, our time response reference model was selected to be

$$\mathbf{M}_0 := \frac{0.12}{s+0.12} I_2, \qquad (4.3)$$

which is fast enough to meet the time response specifications. The loop shaping weighting function we decided on was

$$\mathbf{W} := \frac{500(s+0.55)}{s(11s+1)} I_2, \qquad (4.4)$$

in which the gain, and pole and zeros locations were arrived at by considering the required loop shape at high- and low-frequency. Integral action is used to boost low-frequency gain. The zero at -0.55 is used to reduce the roll-off to approximately 20 dB decade^{-1} at cross-over, while the pole at -0.0909 ensures a low controller bandwidth. Notice that because the two diagonal elements of **W** are the same, **W** is well-conditioned. Ill-conditioned compensators for ill-conditioned plants can give very poor robustness at other points in the loop (Freudenberg, 1990). The corresponding shaped and unshaped open loop singular values plots are given in Fig. 4(a). When selecting the loop shaping function, we were aiming at high low-frequency gain and a low bandwidth controller.

The loop shaping function in (4.4) gives $\gamma_{opt} = 6.2350$ for the pure robustness problem associated with **GW**; see equation (2.10). The design may now be completed by selecting a value for ρ. We found that $\rho = 1.1$ gave a good compromise between acceptable stability properties and meeting the time domain requirements of the design. This value of ρ leads to $\gamma = 8.0105$ for the lowest achievable value of γ in (3.34). For the optimal value of γ, Figs 4(b) and (c) show the Bode magnitude plots for the controller. Both the prefilter and the feedback controller roll off at high frequencies. This gives an acceptably low controller bandwidth.

Figures 5(a) and (b) show closed loop step responses for the inputs $H(t)\begin{bmatrix}1\\0\end{bmatrix}$ and $H(t)\begin{bmatrix}0.4\\0.6\end{bmatrix}$, respectively, when $\tau_1 = \tau_2 = 1.0$ minute, and when $k_1 = k_2 = 1.0$, $k_1 = k_2 = 0.8$, $k_1 = k_2 = 1.2$, $k_1 = 0.8$ and $k_2 = 1.2$ and $k_1 = 1.2$ and $k_2 = 0.8$. The "+" curves illustrate the step responses of the target model given in (4.3). These diagrams show that the closed loop system has a response which is close to the model response for an extreme range of actuator gains, and that the time response specifications are met for the maximum actuator delays of $\tau_1 = \tau_2 = 1.0$

† $H(t)$ is the unit step applied at $t = 0$.

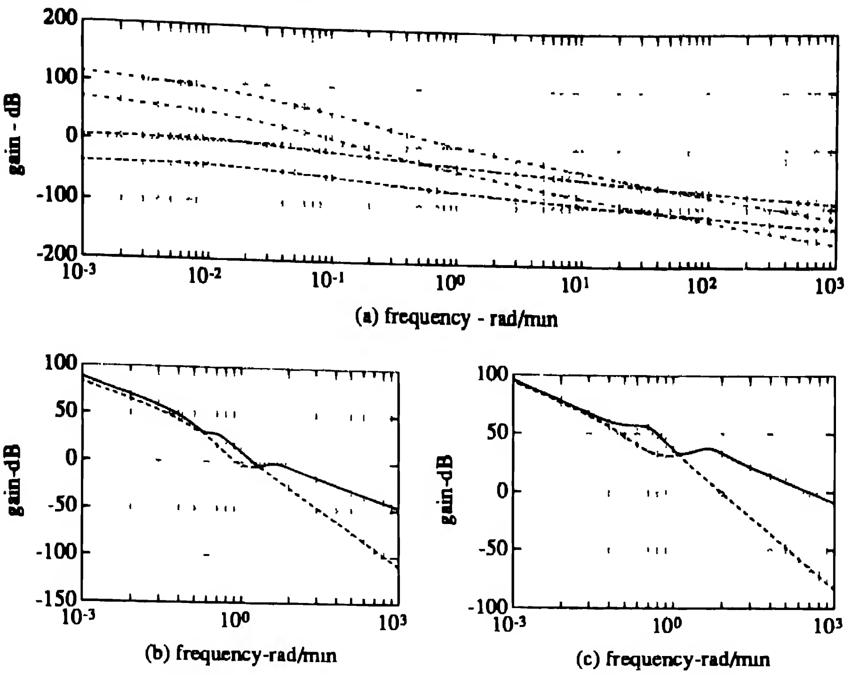


FIG 4 (a) The shaped and unshaped open loop singular values for the LV configuration The prefilter (b) and feedback controller (c) open loop frequency responses

minute. We complete the design study with Fig. 6, which shows the closed loop step responses to the input $H(t)\begin{bmatrix}1\\0\end{bmatrix}$ for nine combinations of τ_1 and τ_2 where τ_1, $\tau_2 = 0.0$, 0.5 and 1.0 minutes, in steps of 0.5 minutes, and every combination of k_1 and k_2 in the range $0.8 \le k_1$, $k_2 \le 1.2$ in steps of 0.2. These curves indicate that the robust performance and robust stability specifications are met for the entire range of allowable time delays and uncertain actuator gains.

4.3. *The two stage design*

The purpose of this section is to evaluate the two step design procedure on the distillation control problem described in Section 4.2. The design study of this section is based on the **G** given in (4.2). As explained earlier and in

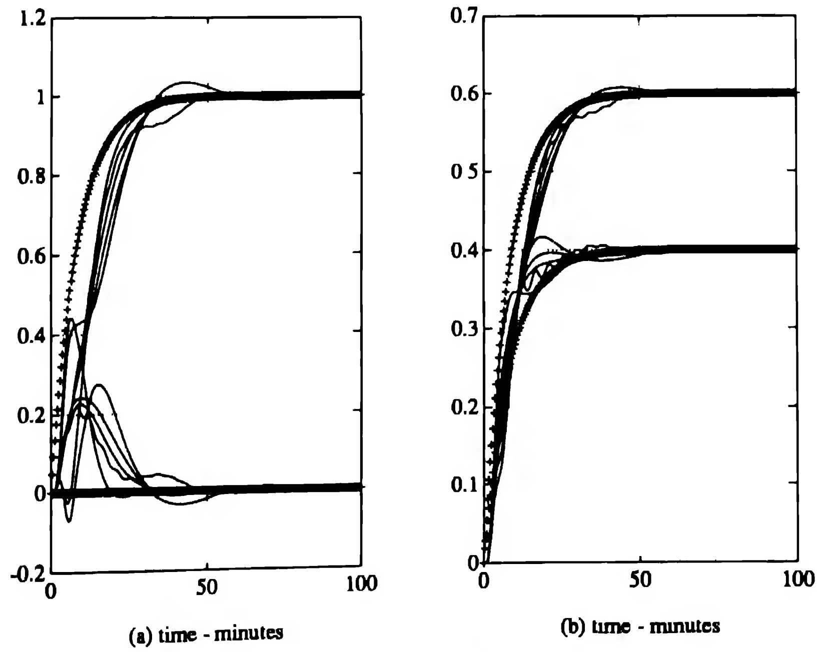


FIG. 5 Closed loop step responses with uncertain actuator gains and nominal time delays of $\tau_1 = \tau_2 = 1.0$ minute.

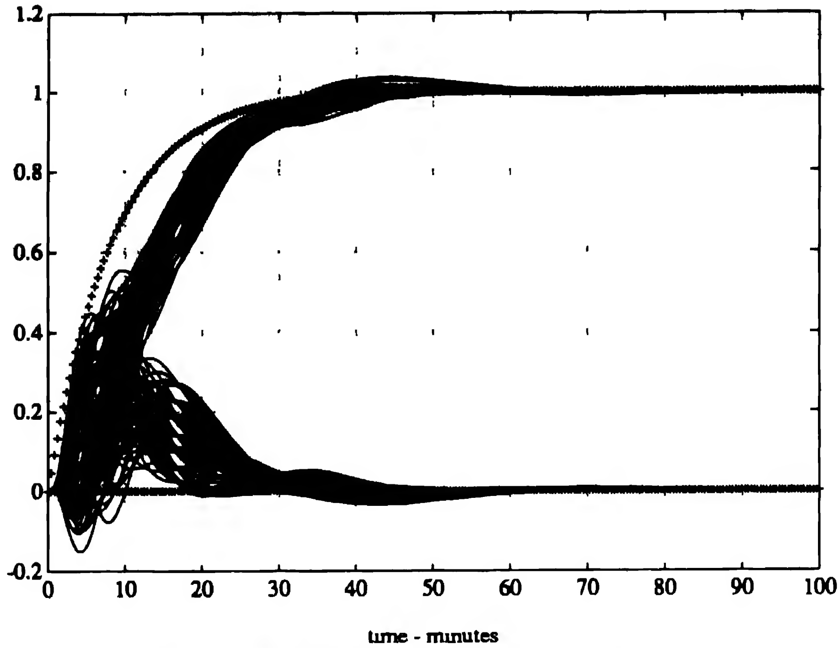


FIG 6 Closed loop step responses with uncertain actuator gains and time delays

McFarlane and Glover (1992) we begin the robust stability optimization by shaping the open loop frequency response. In the present case, we decided on

$$\mathbf{W} = \begin{bmatrix} 1200 & 0 \\ 0 & 260 \end{bmatrix} \frac{(s+0.55)(0.8s+1)}{s(20s+1)(0.1s+1)}, \quad (4.5)$$

as the loop shaping function. In our iterative design trials we found that a weight with a condition number greater than one could be used to improve the performance properties o the closed loop. The $\begin{bmatrix} 1200 & 0 \\ 0 & 260 \end{bmatrix}$ part of th weight was used to marginally reduce th condition number of **GW**. We see from Fig 7(a), which gives the shaped and unshapec frequency responses, that the effect of **W** is to:

(1) Reduce the open loop condition number.
(2) Increase the low-frequency gain. This effec

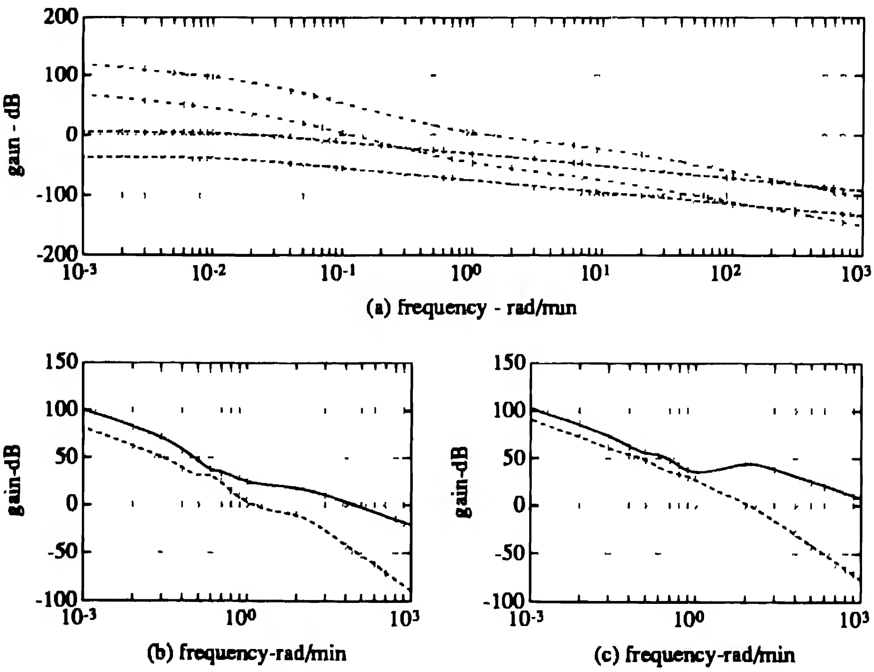


FIG 7 (a) The shaped and unshaped open loop singular values for the LV configuration The prefilter (a) and the feedback controller (c) open loop frequency responses.

is obviously produced by the integrators in the weighting function.

(3) The zeros are $-0.55\,\text{rad minute}^{-1}$ and $-1.2\,\text{rad minute}^{-1}$ and are used to decrease the unity gain roll off rate. Restricting the roll-off rate near unity gain in this way has a beneficial effect on the closed loop time response.

(4) The pole at -0.05 is used to ensure a low controller bandwidth, while the pole at -10 is used to increase the controller's high frequency roll-off rate.

It follows from (2 10) that $\gamma_{opt} = 5.0435$ for $\mathbf{GW}$ when the two unity time delays are approximated by the first order Padé approximation. The feedback controller $\mathbf{K}_2$ was evaluated by substituting $\gamma_2 = \gamma_{opt}$ into equations (2.11) and (2.12). Since $5.0435 < 6.2350$, the two stage design method offers an improved robust stability margin (as compared with the single step procedure). The prefilter was designed by solving for a $\mathbf{Q}_1 \in \mathcal{H}_\infty^+$ which minimizes

$$\gamma_1 = \left\| \begin{bmatrix} \mathbf{M}_0 + \tilde{\mathbf{N}}\mathbf{Q}_1 \\ \tilde{\mathbf{M}}\mathbf{Q}_1 \end{bmatrix} \right\|_\infty . \tag{4.6}$$

The target model $\mathbf{M}_0$ is given in (4.3). A γ-iteration (Glover *et al.*, 1991) shows that the minimum achievable value of γ_1 is $\gamma_1 = 0.6523$. Since

$$\begin{aligned}\sqrt{\gamma_1^2 + \gamma_2^2} &= \sqrt{0.6523^2 + 5.0435^2} \\ &= 5.0855 < 6.2350 \times (1.1)^{-2} = 5.1529,\end{aligned} \tag{4.7}$$

the two stage design gives an improved margin of robust model matching in the face of multiplicative input uncertainty. Figures 7(b) and (c) show the singular value frequency responses for the prefilter and feedback compensators, respectively. The two controllers exhibit high gain at low-frequency which is provided by the integrators in $\mathbf{W}$. Since the controller gain rolls off (at $20\,\text{dB decade}^{-1}$) from steady state, both compensators have acceptably low bandwidths. Figures 8(a) and (b) give closed loop responses for the inputs

$$H(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad H(t) = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix},$$

respectively, when $\tau_1 = \tau_2 = 1.0$ minute and when $k_1 = k_2 = 1.0$, $k_1 = k_2 = 0.8$, $k_1 = k_2 = 1.2$, $k_1 = 0.8$ and $k_2 = 1.2$, and $k_1 = 1.2$ with $k_2 = 0.8$. As a final simulation study, we computed a large set of step responses corresponding to τ_1 and τ_2 in the range of $0 \le \tau_1,\ \tau_2 \le 1$ minute, and k_1 and k_2 in the range $0.8 \le k_1,\ k_2 \le 1.2$. To conserve space, these curves are not illustrated, but they are reminiscent of the results in Fig. 6.

5. CONCLUSIONS

The aim of this paper is to expand the utility of $\mathcal{H}_\infty$ optimization procedures in the design of TDF control systems. We present two new design techniques which are based on the TDF parametrization theory of Youla and Bongiorno (1985) and the loop shaping ideas presented in McFarlane and Glover (1992). A key requirement

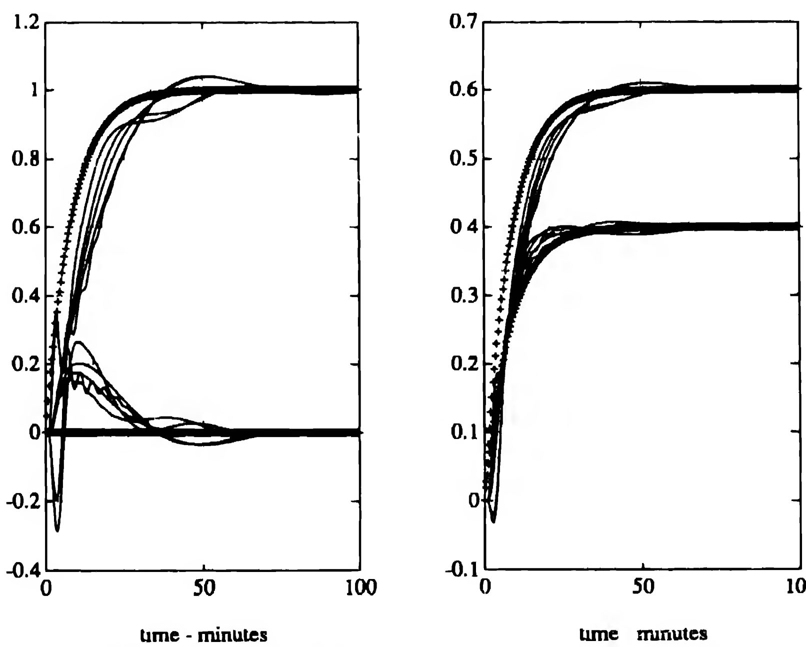


FIG. 8 Closed loop step responses with uncertain actuator gains and nominal time delays of $\tau_1 = \tau_2 = 1.0$ minutes

of both methods is one of robust stability optimization in the face of normalized coprime factor perturbations. Georgiou and Smith have shown that coprime factor robustness is closely related to the idea of robustness in the gap metric (Georgiou and Smith, 1990). These authors are currently developing a TDF design procedure which is based on their gap metric ideas (Georgiou and Smith, 1991).

In order to explicitly address the problem of meeting closed loop time response requirements, our design methods seek to minimize or at least bound quantities of the form $\|(\mathbf{I}-\mathbf{G}\mathbf{K}_2)^{-1}\mathbf{G}\mathbf{K}_1-\mathbf{M}_0\|_\infty$ in which the target $\mathbf{M}_0$ has idealized time response properties. This gives a nominal level of model matching. If the minimization process is expanded to include other terms, one can bound $\|(\mathbf{I}-\mathbf{G}_p\mathbf{K}_2)^{-1}\mathbf{G}_p\mathbf{K}_1-\mathbf{M}_0\|_\infty$ for all plant models $\mathbf{G}_p$ in some model set $\mathcal{G}$. These ideas are explained in Section 2.2 of the paper where we consider model sets generated by perturbed normalized coprime factors and multiplicative perturbations at the plant input. There is no difficulty associated with modifying these ideas to include additive perturbations and multiplicative perturbations at the plant output. Our design procedures are given in Section 3.

Section 4 describes a comprehensive design study for a high purity distillation system. This work shows that our new design procedures may be used to meet a demanding mixture of robust stability and robust performance specifications. The TDF techniques have proved to be particularly useful in problems with explicit time response specifications which have to be met for a wide range of plant parameters.

REFERENCES

Freudenberg, J. S. (1990). Plant directionality, coupling and multivariable loop-shaping. *Int. J. Control*, **51**, 365–390.

Georgiou, T. T. and M. C. Smith (1991). Notes on two degree-of-freedom design. Private communication.

Georgiou, T. T. and M. C. Smith (1990). Optimal robustness in the gap metric. *IEEE Trans. Aut. Control*, **35**, 673–686.

Glover, K., D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally and M. G. Safonov (1991). A characterisation of all solutions to the four block general distance problem. *SIAM J. Control and Optimization*, **29**, 283–324.

Glover, K. and D. C. McFarlane (1989). Robust stabilisation of normalised coprime factor plant descriptions with $\mathcal{H}_\infty$ bounded uncertainty. *IEEE Trans. Aut. Control*, **34**, 821–830.

Limebeer, D. J. N. (1991). The specification and purpose of a controller design case study. *Proc. of the Conf. on Decision and Control*, Brighton, 1579–1580.

McFarlane, D. C. and K. Glover (1992). Robust controller design using normalised coprime factor plant descriptions. *IEEE Trans. Aut. Control*, **37**, 759–769.

Morari, Z. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall, Englewood Cliffs, NJ.

Redheffer, R. M. (1960). On a certain linear fractional transformation. *J. Maths. Phys.*, **39**, 269–286.

Skogestad, S., M. Morari and J. C. Doyle (1988). Robust control of ill-conditioned plants: high-purity distillation. *IEEE Trans. Aut. Control*, **33**, 1092–1105; (1989), μ-optimal controller. *IEEE Trans. Aut. Control*, **34**, 672.

Vidyasagar, M. (1985). *Control System Synthesis—A Factorisation Approach*. MIT Press, MA.

Yaniv, O. and N. Barlev (1990). Robust non-iterative synthesis of ill-conditioned plants. American Control Conference, San Diego.

Youla, D. C. and J. J. Bongiorno (1985). A feedback theory of two-degree of freedom optimal Wiener–Hopf design. *IEEE Trans. Aut. Control*, **30**, 652–665.

Youla, D. C., H. A. Jabr and J. J. Bongiorno (1976). Modern Wiener–Hopf design of optimal controllers—Part II: the multivariable case. *IEEE Trans. Aut. Control*, **21**, 319–338.

Automatica, Vol. 29, No 1, pp 169–179, 1993
Printed in Great Britain

0005–1098/93 $6 00 + 0.00

Robust Tracking and Performance for Multivariable Systems Under Physical Parameter Uncertainties*†

OSVALDO MARIA GRASSELLI,‡ SAURO LONGHI§ and ANTONIO TORNAMBÈ‡

A design method is given for obtaining decoupling, stability, asymptotic tracking and disturbance rejection at the nominal parameters of the plant, and to maintain the last three items in spite of variations and/or uncertainties of some "physical" parameters.

Key Words—Multivariable control systems; robust control, tracking systems; disturbance rejection.

Abstract—In this paper the robust asymptotic tracking and disturbance rejection problem is solved for linear time-invariant systems whose matrices are assumed to depend on some parameters, each of which possibly affects all the elements of the matrices describing the system, thus playing the role of a "physical" parameter It is assumed that reference commands exist only for some of the controlled outputs (i e. that some scalar outputs must only be regulated). For such outputs the row by row decoupling at the nominal parameters is also obtained Both the conditions for the existence of a solution and a design procedure of the compensator are given, the latter enabling us to satisfy some performance requirement in some (possibly "large") subset of the parameter space (with the help of the existing robust stabilization procedures)

1 INTRODUCTION

THE PROBLEM OF THE asymptotic tracking and disturbance rejection of a linear multivariable system subject to unmeasurable disturbances was studied by many authors (see e.g. Basile and Marro, 1991; Basile *et al.*, 1987; Davison, 1976; Davison and Goldenberg, 1975; Desoer and Wang, 1978; Francis, 1977; Francis and Wonham, 1975a, b, 1976; Grasselli and Nicolò, 1971, 1973a, b; Staats and Pearson, 1977; Wonham, 1974; Young and Willems, 1972) and contributions were given even for the case of periodic systems (Grasselli *et al.*, 1979; Colaneri, 1990; Grasselli and Longhi, 1991b). In most of these contributions parameter uncertainties were taken into account, consisting of small (or possibly large) independent perturbations of all the elements of matrices describing the system: it is required the compensator to robustly maintain stability, asymptotic tracking and output regulation in spite of them; in most cases it is assumed that all the controlled outputs must track corresponding reference commands. In the last decade, an increasing number of contributions appeared on the problem of robustly maintaining the asymptotic stability (and possibly, the fulfillment of some performance requirements) for "large" and possibly prescribed perturbations of the parameters of the system, described in different forms (Ackermann, 1980; Barmish, 1985; Bernstein and Haddad, 1990; Doyle and Stein, 1981; Dorato, 1987; Francis, 1987; Galimidi and Barmish, 1986; Gu *et al.*, 1991; Khargonekar *et al.*, 1990; Keel *et al.*, 1988; Petersen and Hollot, 1986; Safonov *et al.*, 1981; Siljak, 1989; Vidyasagar and Kimura, 1986; Wei and Yedavalli, 1989; Willems and Willems, 1983; Zhou and Khargonekar, 1988; and the references therein).

More recently, the problem of the output regulation under perturbations of "physical" parameters affecting the description of the system was solved (Grasselli and Longhi, 1991a). It was shown that robust solutions may exist even when no solution exists for wholly independent variations of the entries of matrices describing the system, e.g. when the physical structure of the plant represented by the system prevents Davison's condition (Davison, 1976;

* Received 6 August 1991; received in final form 26 February 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor P. Dorato

† This work was supported by Ministero Pubblica Istruzione, C.N.R and Ministero Università Ricerca Scientifica Tecnologica

‡ Dipartimento di Ingegneria Elettronica, Seconda Università di Roma "Tor Vergata", via Orazio Raimondo, 00173 Roma, Italy.

§ Dipartimento di Elettronica e Automatica, Università di Ancona, via Brecce Bianche, 60131 Ancona, Italy.

Davison and Goldenberg, 1975) from being satisfied. The aim of this contribution is two-fold: first to extend the results of Grasselli and Longhi (1991a) (with the same kind of parameter dependence) to the asymptotic tracking of some reference signals, possibly for a proper subvector $\bar{y}(t)$ of the controlled output vector $y(t)$ (i.e. it is assumed here that for some components of $y(t)$ there may be no reference signal to be tracked) for the case in which the above-mentioned condition in Davison (1976) and Davison and Goldenberg (1975) is not satisfied; second, to add to the above problem a requirement of decoupling at the nominal parameters (together with some classical SISO requirements on the transient behaviour) and some global performance requirement about $\bar{y}(t)$ to be robustly satisfied in a (possibly "large") subset of the parameter space, with the help of the existing robust stabilization design techniques. In Section 2 the results in Grasselli and Longhi (1991a) are extended to the first problem, and to the case when some other measurable outputs are available in addition to $y(t)$. In Section 3 some performance requirement is added to the same problem, and a design procedure is presented in order to give a solution to the second problem.

2. ROBUST OUTPUT REGULATION AND TRACKING UNDER UNCERTAINTIES OF PHYSICAL PARAMETERS

Consider the linear time-invariant system S described by

$$\dot{x}(t) = A(\beta)x(t) + B(\beta)u(t) + \sum_{i=1}^{\mu} M_i(\beta)d_i(t), \tag{2.1a}$$

$$y(t) = C(\beta)x(t) + D(\beta)u(t) + \sum_{i=1}^{\mu} N_i(\beta)d_i(t), \tag{2.1b}$$

$$\bar{y}(t) = \bar{C}(\beta)x(t) + \bar{D}(\beta)u(t) + \sum_{i=1}^{\mu} \bar{N}_i(\beta)d_i(t), \tag{2.1c}$$

where $t \in \mathbb{R}$ is time, $x(t) \in \mathbb{R}^n =: X$ is the state, $u(t) \in \mathbb{R}^p =: U$ is the control input, $d_i(t) \in \mathbb{R}^{m_i}$ $(i = 1, \ldots, \mu)$ are the unmeasurable and unknown disturbance inputs, $y(t) \in \mathbb{R}^q =: Y$ is the output to be controlled (which is assumed to be measurable), $\bar{y}(t) \in \mathbb{R}^{\bar{q}}$ is the vector of the additional measurable outputs and $A(\beta)$, $B(\beta)$, $C(\beta)$, $D(\beta)$, $\bar{C}(\beta)$, $\bar{D}(\beta)$, $M_i(\beta)$, $N_i(\beta)$, $\bar{N}_i(\beta)$ $(i = 1, \ldots, \mu)$ are matrices with real entries depending on a vector β of parameters, which are subject to variations and/or uncertain, $\beta \in \Omega \subset \mathbb{R}^h$. The dependence on β in (2.1) seems to be able to describe all kinds of uncertainties in modelling a linear time-invariant process through state-space equations, e.g. imperfect knowledge, or perturbations, of "physical" parameters. The nominal value β_0 of β is assumed to be an interior point of the set Ω which is assumed to be bounded, and the values of $A(\beta_0)$, $B(\beta_0)$, $C(\beta_0)$, $D(\beta_0)$, $\bar{C}(\beta_0)$, $\bar{D}(\beta_0)$, $M_i(\beta_0)$, $N_i(\beta_0)$, $\bar{N}_i(\beta_0)$ $(i = 1, \ldots, \mu)$ will be denoted simply by A, B, C, D, $\bar{C}$, $\bar{D}$, M_i, N_i, $\bar{N}_i$ $(i = 1, \ldots, \mu)$. It is assumed that each of the first $\bar{q}$ components $y_1(t), \ldots, y_{\bar{q}}(t)$ of $y(t)$ must track the corresponding component of the reference vector $r(t) \in \mathbb{R}^{\bar{q}}$, $\bar{q} \leq q$. Therefore the error signal $e(t) \in \mathbb{R}^q$ for S is defined by

$$e(t) := Vr(t) - y(t), \quad V := [I \quad 0]'. \tag{2.2}$$

It is also assumed that the classes $\mathcal{R}$ of reference signals $r(\cdot)$ to be asymptotically tracked and $\mathcal{D}_i$ of disturbance functions $d_i(\cdot)$ $(i = 1, \ldots, \mu)$ to be asymptotically rejected are of sinusoidal–exponential type and defined as follows:

$$\mathcal{R} := \mathcal{R}_1 \oplus \mathcal{R}_2 \oplus \cdots \oplus \mathcal{R}_{\bar{\mu}}, \tag{2.3a}$$

$$\mathcal{R}_i := \{r(\cdot): r(t) = \delta e^{\alpha_i t} + \delta^* e^{\alpha_i^* t}\ \forall t \geq 0,\ \delta \in \mathbb{C}^{\bar{q}}\} \quad i = 1, \ldots, \bar{\mu}, \tag{2.3b}$$

$$\mathcal{D}_i := \{d_i(\cdot): d_i(t) = \delta e^{\alpha_i t} + \delta^* e^{\alpha_i^* t}\ \forall t \geq 0,\ \delta \in \mathbb{C}^{m_i}\} \quad i = 1, \ldots, \mu, \tag{2.3c}$$

for some positive integer $\bar{\mu} \leq \mu$ and some $\alpha_i \in \mathbb{C}$ $(i = 1, \ldots, \mu)$, where * means complex conjugate. Notice that the triplets $(M_i(\cdot), N_i(\cdot), \bar{N}_i(\cdot))$ and $(M_j(\cdot), N_j(\cdot), \bar{N}_j(\cdot))$ may actually coincide for some $i, j, i \neq j$. In order to rule out trivialities, the α_is $(i = 1, \ldots, \mu)$ are assumed to be all distinct, and to have all non-negative real and imaginary parts, while the triplet $(M_i(\cdot), N_i(\cdot), \bar{N}_i(\cdot))$ is assumed to be non-zero for each $i = \bar{\mu} + 1, \ldots, \mu$.

The problem of designing a control system for S so that suitable requirements are satisfied, will be studied under the following technical assumption.

Assumption 1. There exists a neighbourhood $\Psi_a \subset \Omega$ of β_0 such that all the entries of $A(\beta)$, $B(\beta)$, $C(\beta)$, $D(\beta)$, $\bar{C}(\beta)$, $\bar{D}(\beta)$ are continuous functions of β in Ψ_a and such that

$$\operatorname{rank}\begin{bmatrix} A(\beta) - \alpha_i I & B(\beta) \\ C(\beta) & D(\beta) \end{bmatrix} = \operatorname{rank}\begin{bmatrix} A - \alpha_i I & B \\ C & D \end{bmatrix} =: \eta_i\ \forall \beta \in \Psi_a, \quad i = 1, \ldots, \mu. \tag{2.4}$$

Lemma 1. If there exists a neighbourhood $\bar{\Psi}_a \subset \Omega$ of some $\bar{\beta}_0 \in \Omega$ such that all the entries of $A(\beta)$, $B(\beta)$, $C(\beta)$, $D(\beta)$, $\bar{C}(\beta)$, $\bar{D}(\beta)$ are

continuous functions of β in Ψ_a, then there exists an interior point β_0 of Ψ_a such that Assumption 1 holds.

The proof of Lemma 1 is reported in the Appendix. By Lemma 1, Assumption 1 is implied by the mere continuity of the matrices involved, provided that the nominal value β_0 of β can be properly chosen adequately near some initial choice $\tilde{\beta}_0$ of it. Therefore, Assumption 1 seems to be reasonable for physical plants.

Denoting with $\hat{y}(t) := [y'(t)\ \ \bar{y}'(t)]'$ the vector of all measured outputs, the linear dynamic compensator K to be designed for S is represented by

$$\dot{w}(t) = Pw(t) + Qr(t) + R\hat{y}(t), \quad (2.5a)$$

$$u(t) = Hw(t) + Jr(t), \quad (2.5b)$$

with $w(t) \in \mathbb{R}^{\nu}$ and P, Q, R, H and J real. Define also

$$\hat{C}(\beta) := [C'(\beta)\ \ \bar{C}'(\beta)]',$$
$$\hat{D}(\beta) := [D'(\beta)\ \ \bar{D}'(\beta)]',$$
$$\hat{N}_i(\beta) := [N_i'(\beta)\ \ \bar{N}_i'(\beta)]',$$
$$i = 1, \ldots, \mu, \quad (2.6)$$

and let $\hat{C}(\beta_0)$, $\hat{D}(\beta_0)$, $\hat{N}_i(\beta_0)$ be denoted simply by $\hat{C}$, $\hat{D}$, $\hat{N}_i$ $(i = 1, \ldots, \mu)$, respectively.

The following *robust tracking and regulation problem* will be considered in this section: find if it exists, a linear dynamic compensator K described by (2.5), such that:

(a) the overall control system Σ described by (2.1) and (2.5), is asymptotically stable at the nominal parameters of system S, i.e. for $\beta = \beta_0$;

(b) the error response $e(t)$ asymptotically goes to zero for each disturbance function $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$ and for each reference signal $r(\cdot) \in \mathcal{R}$, for $\beta = \beta_0$;

(c) properties (a) and (b) are preserved for all β in some neighbourhood $\Psi \subset \Omega$ of β_0 (or possibly in some "large" subset Ψ of Ω containing β_0 as an interior point).

In the following, such a problem will be called *Problem* 1, and property (b) will be called *asymptotic tracking and disturbance rejection*.

Theorem 1. There exists a solution K of Problem 1, under Assumption 1, if and only if:

(i) the triplet $(A, B, \hat{C})$ is stabilizable and detectable; and,

(ii) there exists a neighbourhood $\Psi_b \subset \Omega$ of β_0 such that

$$\mathrm{Im}\begin{bmatrix} M_i(\beta) \\ N_i(\beta) \end{bmatrix} \subset \mathrm{Im}\begin{bmatrix} A(\beta) - \alpha_i I & B(\beta) \\ C(\beta) & D(\beta) \end{bmatrix}$$
$$\forall \beta \in \Psi_b \quad i = 1, \ldots, \mu, \quad (2.7a)$$

$$\mathrm{Im}\begin{bmatrix} 0 \\ V \end{bmatrix} \subset \mathrm{Im}\begin{bmatrix} A(\beta) - \alpha_i I & B(\beta) \\ C(\beta) & D(\beta) \end{bmatrix}$$
$$\forall \beta \in \Psi_b \quad i = 1, \ldots, \bar{\mu}. \quad (2.7b)$$

Proof. (Necessity) Property (a) trivially implies condition (i). By virtue of (a), (b) and (c), the steady-state responses of Σ to the disturbance functions $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$ and reference signals $r(\cdot) \in \mathcal{R}$ written for $x(t)$, $u(t)$ and $e(t)$, imply condition (ii).

(Sufficiency) Define

$$\rho_i := \eta_i - n, \quad i = 1, \ldots, \mu, \quad (2.8)$$

where $\rho_i \geq 0$ by the stabilizability of (A, B) and the assumptions about α_i $(i = 1, \ldots, \mu)$.

Condition (i) implies that system S can be stabilized at the nominal parameters by a linear dynamic feedback from the measured output $\hat{y}(t)$. Let such a stabilization be obtained through a dynamic control law of the following form:

$$\dot{\xi}(t) = (A - L\hat{C})\xi(t) + (B - L\hat{D})u(t) + L\hat{y}(t),$$
$$\xi(t) \in \mathbb{R}^n, \quad (2.9a)$$

$$u(t) = F\xi(t) + u_c(t), \quad u_c(t) \in \mathbb{R}^p, \quad (2.9b)$$

i.e. feeding back through a constant feedback gain matrix F the state $\xi(t)$ of an estimator of $x(t)$ for $d_i(\cdot) = 0$, $i = 1, \ldots, \mu$, in which, by virtue of (i), the eigenvalues of $(A - L\hat{C})$ are obtained to lie in the open left half-plane, as well as those of $(A + BF)$. Then, it is easily seen that the closed-loop system S_c described by (2.1) and (2.9) of dimension $n_c := 2n$, having $u_c(t)$ as control input and $y(t)$ as the only measured and controlled output, and described by equations similar to (2.1a) and (2.1b) but involving $u_c(t)$ as control input instead of $u(t)$ and some matrices $A_c(\beta)$, $B_c(\beta)$, $C_c(\beta)$, $D_c(\beta)$, $M_{ci}(\beta)$, $N_{ci}(\beta)$ $(i = 1, \ldots, \mu)$ instead of the corresponding matrices in (2.1a) and (2.1b), satisfies the following condition (where a notation similar to the corresponding one defined for S is used for the matrices describing S_c at $\beta = \beta_0$).

Condition 1. Assumption 1 and condition (ii) of the theorem, rewritten for the matrices describing S_c, hold for the same neighbourhoods Ψ_a, Ψ_b, the triplet (A_c, B_c, C_c) is stabilizable and detectable, and the following relation holds:

$$\mathrm{rank}\begin{bmatrix} A_c - \alpha_i I & B_c \\ C_c & D_c \end{bmatrix} - n_c = \rho_i, \quad i = 1, \ldots, \mu. \quad (2.10)$$

If, in particular, $\bar{y}(t)$ vanishes (i.e. no additional measurement is available) and/or the pair (A, C) is detectable (i.e. no additional measurements $\bar{y}(t)$ are needed in order to

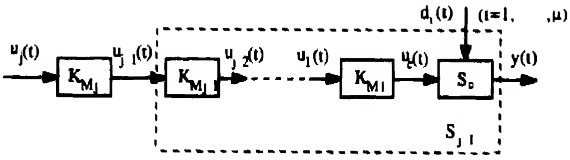


FIG. 1. The series connection S_j of K_{M_j}, $K_{M_{j-1}}, \ldots, K_{M_1}$ and S_c.

stabilize system S) the dynamic control law (2.9) is not yet needed and can vanish, so that S_c can coincide with S, while preserving Condition 1.

It will now be shown the existence of μ sub-compensators $K_{M_1}, K_{M_2}, \ldots, K_{M_\mu}$ such that, for each $j = 1, \ldots, \mu$, the series connection S_j of $K_{M_j}, K_{M_{j-1}}, \ldots, K_{M_1}$ and S_c, having the input $u_j(t) \in \mathbb{R}^p$ of K_{M_j} as the control input and $y(t)$ as output (see Fig. 1) and described by

$$\dot{x}_j(t) = A_j(\beta)x_j(t) + B_j(\beta)u_j(t) + \sum_{i=1}^{\mu} M_{ji}(\beta)d_i(t), \quad x_j(t) \in \mathbb{R}^{n_j}, \quad u_j(t) \in \mathbb{R}^p, \tag{2.11a}$$

$$y(t) = C_j(\beta)x_j(t) + D_j(\beta)u_j(t) + \sum_{i=1}^{\mu} N_{ji}(\beta)d_i(t), \tag{2.11b}$$

satisfies the following conditions.

Condition 2. The above stated Condition 1, rewritten for system S_j, its dimension n_j and the matrices describing it, hold.

Condition 3. There exists a neighbourhood $\Psi_j \subset \Omega$ of β_0 such that

$$\operatorname{Im}\begin{bmatrix} M_{jj}(\beta) \\ N_{jj}(\beta) \end{bmatrix} + \operatorname{Im}\begin{bmatrix} 0 \\ \theta_j V \end{bmatrix} \subset \operatorname{Im}\begin{bmatrix} A_j(\beta) - \alpha_j I \\ C_j(\beta) \end{bmatrix} \quad \forall \beta \in \Psi_j, \tag{2.12a}$$

$$\theta_j := 1 \quad \text{if } j \le \bar{\mu}, \quad \theta_j := 0 \quad \text{if } j > \bar{\mu}. \tag{2.12b}$$

The proof of the existence of $K_{M_1}, \ldots, K_{M_\mu}$ with the above stated properties will be carried out constructively by induction. Then, assume that S_{j-1} satisfies Condition 2 written for S_{j-1} (for $j = 1$ this is true if S_c is denoted with S_0 and $u_c(t)$ with $u_0(t)$). If $\alpha_j \in \mathbb{R}$, let K_{M_j} be described by

$$\dot{w}_j(t) = \alpha_j w_j(t) + [0 \quad I]u_j(t), \quad w_j(t) \in \mathbb{R}^{\rho_j}, \quad u_j(t) \in \mathbb{R}^p, \tag{2.13a}$$

$$u_{j-1}(t) = E_j w_j(t) + [\hat{E}_j \quad 0]u_j(t), \quad u_{j-1}(t) \in \mathbb{R}^p, \tag{2.13b}$$

where $\rho_j \le p$ by (2.4) and (2.8), $\hat{E}_j \in \mathbb{R}^{p \times (p-\rho_j)}$, E_j and $\hat{E}_j$ are real, and E_j and $\hat{E}_j$ are such that, using for the matrices describing S_{j-1} at $\beta = \beta_0$ a notation similar to the corresponding one defined for S,

$$\operatorname{rank}\begin{bmatrix} A_{j-1} - \alpha_j I & B_{j-1}E_j \\ C_{j-1} & D_{j-1}E_j \end{bmatrix} = n_{j-1} + \rho_j \quad \text{(i.e. full column-rank)}, \tag{2.14}$$

$$\det[E_j \quad \hat{E}_j] \neq 0. \tag{2.15}$$

If $\alpha_j \notin \mathbb{R}$, define through (2.14) and (2.15) the real matrices E_j and $\hat{E}_j$ in the same way (and with the same dimensions) as for $\alpha_j \in \mathbb{R}$, and let K_{M_j} be described by

$$\dot{w}_j(t) = W_j w_j(t) + [0 \quad \bar{U}_j]u_j(t), \quad w_j(t) \in \mathbb{R}^{2\rho_j}, \quad u_j(t) \in \mathbb{R}^p, \tag{2.16a}$$

$$u_{j-1}(t) = \bar{E}_j w_j(t) + [\hat{E}_j \quad 0]u_j(t), \quad u_{j-1}(t) \in \mathbb{R}^p, \tag{2.16b}$$

where $\bar{E}_j$ is defined by

$$\bar{E}_j = [e_j^1 \quad 0 \quad e_j^2 \quad 0 \cdots e_j^{\rho_j} \quad 0], \tag{2.17}$$

in which e_j^i represents the ith column of E_j ($i = 1, \ldots, \rho_j$), W_j is defined by

$$W_j = \operatorname{diag}\left\{\begin{bmatrix} \sigma_j & \omega_j \\ -\omega_j & \sigma_j \end{bmatrix}, \ldots, \begin{bmatrix} \sigma_j & \omega_j \\ -\omega_j & \sigma_j \end{bmatrix}\right\}, \tag{2.18}$$

in which σ_j and ω_j are the real and the imaginary parts, respectively, of α_j (with $\omega_j \neq 0$), and $\bar{U}_j \in \mathbb{R}^{2\rho_j \times \rho_j}$ is defined by

$$\bar{U}_j = \operatorname{diag}\left\{\begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}. \tag{2.19}$$

Now, for both types (2.13) and (2.16) of subcompensator K_{M_j}, it can be easily seen that (2.14), (2.15) and the assumption that S_{j-1} satisfies Condition 2 (written for it) for the neighbourhoods $\Psi_a, \Psi_b \subset \Omega$ of β_0 imply that S_j satisfies Condition 2 for the same neighbourhoods Ψ_a, Ψ_b. In addition, for both types (2.13) and (2.16) of K_{M_j}, condition (2.12a) holds if and only if

$$\operatorname{Im}\begin{bmatrix} M_{j-1,j}(\beta) \\ N_{j-1,j}(\beta) \end{bmatrix} + \operatorname{Im}\begin{bmatrix} 0 \\ \theta_j V \end{bmatrix} \subset \operatorname{Im}\begin{bmatrix} A_{j-1}(\beta) - \alpha_j I & B_{j-1}(\beta)E_j \\ C_{j-1}(\beta) & D_{j-1}(\beta)E_j \end{bmatrix} \quad \forall \beta \in \Psi_j. \tag{2.20}$$

On the other hand, Assumption 1 on S_{j-1} and (2.14) imply the existence of some neighbourhood $\Psi_{0j} \subset \Psi_a$ of β_0 in which the dimension of the subspace in the right-hand side of (2.20) remains constant, whence by (2.4) rewritten for

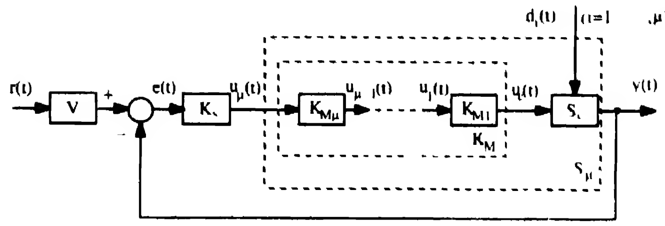


FIG 2 The overall control system Σ for Problem 1

S_{j-1} and (2.14), such that

$$\operatorname{Im}\begin{bmatrix} A_{j-1}(\beta)-\alpha_j I & B_{j-1}(\beta)E_j \\ C_{j-1}(\beta) & D_{j-1}(\beta)E_j \end{bmatrix} = \operatorname{Im}\begin{bmatrix} A_{j-1}(\beta)-\alpha_j I & B_{j-1}(\beta) \\ C_{j-1}(\beta) & D_{j-1}(\beta) \end{bmatrix} \quad \forall \beta \in \Psi_{0j}. \tag{2.21}$$

This, together with (2.7) rewritten for S_{j-1}, proves (2.20) with $\Psi_j = \Psi_{0j} \cap \Psi_b$, and completes the inductive proof that, for each $j = 1, \dots, \mu$, S_j satisfies Conditions 2 and 3.

Since, by Condition 2, S_μ is stabilizable and detectable at the nominal parameters β_0, there exists for it a linear compensator K_s described by

$$\dot{w}_s(t) = P_s w_s(t) + Q_s e(t), \tag{2.22a}$$

$$u_\mu(t) = H_s w_s(t), \tag{2.22b}$$

with P_s, Q_s and H_s real, such that the overall contro' system Σ thus obtained, represented in Fig. 2, satisfies requirement (a).

Denote by $A_\Sigma(\beta)$ the dynamic matrix of Σ. Since, by virtue of Assumption 1, the coefficients of the characteristic polynomial of $A_\Sigma(\beta)$ are continuous functions of β in Ψ_a, there certainly exists a neighbourhood $\Psi_c \subset \Omega$ of β_0 such that the asymptotic stability of Σ is preserved for all $\beta \in \Psi_c$. This, together with (2.12) for each $j = 1, \dots, \mu$, implies that asymptotic tracking and disturbance rejection are guaranteed for Σ at $\beta = \beta_0$ and for all $\beta \in \Psi_c \cap \Psi_1 \cap \Psi_2 \cap \dots \cap \Psi_\mu =: \Psi$, by virtue of Theorem 1 in Grasselli and Nicolò (1976).

Remark 2.1. If the stabilizing compensator K_s (and possibly, F and L in (2.9)) is designed according to any procedure which is able to guarantee the asymptotic stability of Σ not only in some neighbourhood $\Psi_c \subset \Omega$ of β_0, but in some "large", and possibly prescribed region Ψ_s of Ω (see e.g. Ackermann, 1980; Barmish, 1985; Bernstein and Haddad, 1990; Doyle and Stein, 1981; Dorato, 1987; Francis, 1987; Galimidi and Barmish, 1986; Gu *et al.*, 1991; Khargonekar *et al.*, 1990; Keel *et al.*, 1988; Peterson and Hollot, 1986; Safonov *et al.*, 1981; Siljak, 1989; Vidyasagar and Kimura, 1986; Wei and Yedavalli, 1989; Willems and Willems, 1983; Zhou and Khargonekar, 1988; and the references therein) and if (2.4) and (2.7) hold for some "large" subsets Ψ_a and Ψ_b of Ω (see Grasselli and Longhi, 1991a, for a simple example of a physical process S for which $\Psi_a = \Psi_b = \Omega$) property (b) too can be preserved in some "large" subset of Ω, and, possibly, in $\Psi_a \cap \Psi_b \cap \Psi_s$. In fact, asymptotic tracking and disturbance rejection are guaranteed for Σ at each value of β such that $A_\Sigma(\beta)$ is asymptotically stable and the following relations hold for S_μ (see Theorem 1 in Grasselli and Nicolò, 1976):

$$\operatorname{Im}\begin{bmatrix} M_{\mu i}(\beta) \\ N_{\mu i}(\beta) \end{bmatrix} + \operatorname{Im}\begin{bmatrix} 0 \\ \theta_i V \end{bmatrix} \subset \operatorname{Im}\begin{bmatrix} A_\mu(\beta)-\alpha_i I \\ C_\mu(\beta) \end{bmatrix}, \quad i = 1, \dots, \mu, \tag{2.23}$$

which can be seen to be equivalent to the following ones, regarding S_i $(i = 1, \dots, \mu)$:

$$\operatorname{Im}\begin{bmatrix} M_{ii}(\beta) \\ N_{ii}(\beta) \end{bmatrix} + \operatorname{Im}\begin{bmatrix} 0 \\ \theta_i V \end{bmatrix} \subset \operatorname{Im}\begin{bmatrix} A_i(\beta)-\alpha_i I \\ C_i(\beta) \end{bmatrix}, \quad i = 1, \dots, \mu, \tag{2.24}$$

(which are guaranteed to hold at each $\beta \in \bigcap_{i=1}^{\mu} \Psi_i \subset \Psi_a \cap \Psi_b$ by the design procedure).

Remark 2.2 The design procedure of K in the sufficiency proof of Theorem 1, valid under Assumption 1 and conditions (i) and (ii), can be summarized as follows.

Step 1. If (A, C) is detectable, denote S with S_0 and use the notations in (2.11) with $j = 0$ for (2.1a) and (2.1b). If (A, C) is not detectable, choose F and L in (2.9) so that the spectra of both $(A + BF)$ and $(A - L\hat{C})$ belong to the open left half-plane; call S_0 the system described by (2.1), (2.6) and (2.9) having $u_c(t)$ as control input and $y(t)$ as output, and use the notations in (2.11) with $j = 0$ for its description.

Step 2. For each $j = 1, \dots, \mu$, if $\alpha_j \in \mathbb{R}$, or $\alpha_j \notin \mathbb{R}$, define K_{M_j} through (2.13) or, respectively

(2.16), with ρ_j defined by (2.4) and (2.8), and E_j, $\bar{E}_j$, $\bar{E}_j$, W_j and $\tilde{U}_j$ satisfying (2.14), (2.15), (2.17), (2.18) and (2.19) for system S_{j-1}; define the notations in (2.11) for the system S_j obtained by connecting K_{M_j} and S_{j-1}.

Step 3. Find P_s, Q_s and H_s for the compensator K_s described by (2.22), so that the overall control system Σ satisfies requirement (a).

At Step 1 (2.9a) can be substituted by any kind of reduced order state estimator of S; this preserves the validity of Condition 1 for S_c. As mentioned in Remark 2.1, robust stabilization design procedures are obviously preferable at both Steps 1 (e.g. the one proposed in Doyle and Stein, 1981) and 3.

It seems worth mentioning that the purpose of Step 2 is to choose K_{M_j}, $j = 1, \ldots, \mu$, so that S_μ, for each β in some neighbourhood $\Psi_d := \bigcap_{i=1}^{\mu} \Psi_i \subset \Psi_a \cap \Psi_b$ of β_0, is able to generate free output responses equal to the output responses of S_c (for the same value of β) to all disturbance functions $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$ and to all signals $Vr(\cdot)$, $r(\cdot) \in \mathcal{R}$ (according to a simple form of the internal model principle (Grasselli and Nicolò, 1971)) while maintaining for S_μ the properties of stabilizability and detectability at $\beta = \beta_0$, enjoyed by system S_c. The former property of S_μ is achieved by attaining (2.24), whence (2.23), to hold for all $\beta \in \Psi_d$. Thus, the meaning of Step 2 is to enlarge (as much as consistent with the stability of Σ) the (partial) internal model of disturbance functions $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$ and signals $Vr(\cdot)$, $r(\cdot) \in \mathcal{R}$, possibly contained in S_c (e.g. an integrator if $S_c = S$ and S contains a motor, and $\alpha_i = 0$ for some i) by means of the series connection of the internal models $K_{M_\mu}, \ldots, K_{M_1}$, so that (2.23) holds for all β in some (possibly large) Ψ_d.

Remark 2.3. Let Ψ_a (or Ψ_b) be the largest subset of Ω such that Assumption 1 (or (2.7)) holds (it is related to the structure of the physical plant). Then a procedure for determining a good estimate $\hat{\Psi}_a \subset \Psi_a$ of Ψ_a can be obtained as follows. Call $\bar{\Psi}_a \supset \Psi_a$ the subset of Ω in which all the elements of $A(\beta)$, $B(\beta)$, $\hat{C}(\beta)$, $\hat{D}(\beta)$ are continuous functions of β. For each $i = 1, \ldots, \mu$, through elementary row and column operations, properly modified in order to avoid multiplications by functions of β which are not continuous in $\bar{\Psi}_a$ or which are null at $\beta = \beta_0$, the argument matrix in the right-hand sides of (2.7) can be transformed into the following form:

$$\begin{bmatrix} \Delta_{1i}(\beta) & 0 \\ 0 & \Delta_{2i}(\beta) \end{bmatrix}, \tag{2.25}$$

where $\Delta_{1i}(\beta)$ is a diagonal matrix whose diagonal elements are continuous functions of β in $\bar{\Psi}_a$ and are not null at $\beta = \beta_0$, by Assumption 1, the elements of $\Delta_{2i}(\beta)$ are continuous functions of β in $\bar{\Psi}_a$, and there exists a subset $\bar{\Psi}_{ai}$ of $\bar{\Psi}_a$, containing β_0 as an interior point, such that $\Delta_{2i}(\beta) = 0$ for all $\beta \in \bar{\Psi}_{ai}$ (by the same Assumption 1). A subset $\hat{\Psi}_{ai}$ can be obtained from $\bar{\Psi}_{ai}$ by eliminating from $\bar{\Psi}_{ai}$ all the values of β where $\det \Delta_{1i}(\beta) = 0$. Then, $\hat{\Psi}_a$ is obtained as $\bigcap_{i=1}^{\mu} \hat{\Psi}_{ai}$. The computation of a good estimate $\hat{\Psi}_b \subset \Psi_b$ of Ψ_b can be similarly obtained by making exactly the same elementary row operations on the argument matrices in the left-hand sides of (2.7).

Remark 2.4. If $\bar{q} = q$, it is easily seen that conditions (i) and (2.7b), written at $\beta = \beta_0$, imply $\rho_i = q$ $(i = 1, \ldots, \bar{\mu})$, i.e. the well-known condition (see e.g. Davison, 1976; Davison and Goldenberg, 1975)

$$\operatorname{rank}\begin{bmatrix} A - \alpha_i I & B \\ C & D \end{bmatrix} = n + q, \quad i = 1, \ldots, \bar{\mu}, \tag{2.26}$$

which, for $\bar{\mu} = \mu$, under Assumption 1, is stronger than the existence of a neighbourhood Ψ_b of β_0 such that (2.7a) holds. However (2.26) is not necessary, in general, if $\bar{q} < q$, and cannot be satisfied if $q > p$, i.e. if system S has more regulated outputs than control inputs. In addition, when $q \leq p$, although condition (2.26) rewritten with μ instead of $\bar{\mu}$ is generically satisfied, i.e. it is satisfied for almost all (A, B, C, D)-tuples (Davison and Wang, 1974), in practice the physical plant described by S may have a physical structure implying the violation of such a condition either because of a non-full row-rank Rosenbrock system matrix of S at $\beta = \beta_0$ and near β_0, or because of one (or more) invariant zero at $\beta = \beta_0$ (and, possibly, near β_0) equal to one of the α_is characterizing $\mathcal{D}_i$ and $\mathcal{R}$. In these cases, as well as when $q > p$, (2.26) written with μ instead of $\bar{\mu}$ is violated for some or for all is, so that a fully redundant internal model of reference signals $r(\cdot) \in \mathcal{R}$ and disturbance functions $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$, put in K (as in Davison, 1976; Davison and Goldenberg, 1975), is not compatible with the stability of Σ. Hence in such cases only a solution K or Problem 1 exists, if Assumption 1 and conditions (i) and (ii) hold, although (2.26) written with μ instead of $\bar{\mu}$ does not hold and the above-mentioned fully redundant internal model of disturbance functions and reference signals cannot be used in K.

Remark 2.5. Statements similar to Theorem 1, with proper amendments of condition (i), hold for the case when requirement (a) is strengthened by prescribing the degree of exponential stability of Σ at $\beta = \beta_0$, or even the spectrum of $A_\Sigma(\beta_0)$ and, correspondingly, requirement (c) is properly tightened, and/or system S is substituted with a discrete-time one, possibly with a dead-beat convergence requirement, as in Grasselli and Longhi (1991a) where the special case $\bar{q} = \bar{q} = 0$ was studied.

3. ADDITIONAL DECOUPLING AND PERFORMANCE REQUIREMENTS

The design procedure reported in the previous section allows to find a solution of Problem 1, even when, for $\bar{q} < q$, (2.26) is not satisfied, since it is able to deal with the kind of parameter dependence assumed here for S, described by (2.1), which can really model the actual dependence of a physical plant on some "physical" parameters. Although this can appear advantageous with respect to other existing methods, it seems not to allow, in general, to satisfy performance requirements on the transient behaviour of Σ.

In the case, usually considered, when (2.26) rewritten with μ instead of $\bar{\mu}$ holds (which is necessary if $\bar{q} = q$ and $\bar{\mu} = \mu$) the so-called row by row decoupling of Σ at $\beta = \beta_0$ can be easily achieved together with requirements (a), (b) and (c) (see e.g. Chen, 1984). In such a case, also some performance requirement can be satisfied, following the approach in Doyle and Stein (1981).

When, on the contrary, $\bar{q} < q$ and condition (2.26) does not hold, a solution to the problem of satisfying all the above mentioned requirements is not available, at the authors' knowledge. In this connection notice that, if $\bar{q} > 0$, the row by row decoupling at the nominal parameters of S, as far as the vector $\bar{y}(t)$ of the first $\bar{q}$ components of $y(t)$ is concerned, can very often be seen as a desirable, although not unrenunciable, feature of Σ. Hence, it seems reasonable to try to satisfy this requirement for Σ together with (a), (b) and (c):

(d) under the assumption $\bar{q} > 0$, the transfer function matrix from $r(t)$ to $\bar{y}(t)$ at the nominal parameters of S (i.e. at $\beta = \beta_0$) denoted here with $W(s)$, is diagonal and nonsingular.

In addition, the following requirements will be considered for Σ:

(e) under the assumption $\bar{q} > 0$, the diagonal elements of $W(j\omega)$ and/or the corresponding scalar step responses satisfy some classical SISO requirements on the transient behaviour, such as bandwidth or risetime, overshoot, etc.,

(f) some performance requirement on Σ, such as the one considered in Doyle and Stein (1981) and specified later on (see the subsequent relation (3.10)) is satisfied for all β in the same set Ψ as in (c).

In the following, Problem 1 with the additional requirements (d), or (d) and (e), or (d), (e) and (f), will be called Problem 2, or 3, or 4, respectively.

Theorem 2. If $\bar{q} > 0$, there exists a solution K of Problem 2, under Assumption 1, if and only if conditions (i) and (ii) of Theorem 1 hold.

Proof. Conditions (i) and (ii) are necessary by Theorem 1. As regards sufficiency, firstly apply the design procedure described in the sufficiency proof of Theorem 1, with the only amendment of substituting $e(t)$ in (2.22a) with $e_0(t) \in \mathbb{R}^{\bar{q}}$, defined by

$$e_0(t) := Vr_0(t) - y(t), \quad r_0(t) \in \mathbb{R}^{\bar{q}}, \qquad (3.1)$$

so that the scheme of the control system in Fig. 2 is modified by the substitution of $e(t)$ and $r(t)$ with $e_0(t)$ and $r_0(t)$, respectively. Call Σ_0, instead of Σ, the control system thus obtained. Call $W_0(s)$ the $\bar{q} \times \bar{q}$ transfer matrix of Σ_0 from $r_0(t)$ to $\bar{y}(t)$ at $\beta = \beta_0$.

Obviously, if $r_0(t) = r(t)$, then $e_0(t) = e(t)$. In this case, by the sufficiency proof of Theorem 1, Σ_0 satisfies requirements (a), (b) and (c); call Ψ_0, instead of Ψ, the neighbourhood of β_0 in which (a) and (b) are preserved. Properties (a) and (b) imply that, for each $i = 1, \ldots, \bar{\mu}$, $W_0(\alpha_i) = W_0(\alpha_i^*) = I$. Therefore, the rational and square matrix $W_0(s)$ has (normal) rank $\bar{q}$ and has no loss of rank at α_i and α_i^*, i.e. no transmission zero at α_i and α_i^* ($i = 1, \ldots, \bar{\mu}$).

Now, let

$$Z(s)T^{-1}(s) = W_0(s), \qquad (3.2)$$

be an irreducible matrix fraction description of $W_0(s)$, i.e. let $Z(s)$ and $T(s)$ be $\bar{q} \times \bar{q}$ right coprime polynomial matrices, with $T(s)$ nonsingular, such that (3.2) holds. By the stability of Σ_0 at $\beta = \beta_0$, $\det T(s)$ has no root in the closed right half-plane, whence the invariant polynomial of $T(s)$ of the smallest degree, denoted here with $\zeta(s)$, is a Hurwitz polynomial. Define $\bar{T}(s) := T(s)/\zeta(s)$. By the nonsingularity of $W_0(s)$, $Z(s)$ is nonsingular; since $W_0(s)$ has no transmission zeros at α_i and α_i^*, $Z(s)$ has no loss of rank at α_i and α_i^* ($i = 1, \ldots, \bar{\mu}$). For each $i = 1, \ldots, \bar{q}$, call $\varepsilon_i(s)$ the monic greatest common divisor of all the elements of the ith row of $Z(s)$, and denote with $\bar{Z}(s)$ the nonsingular polynomial matrix obtained from $Z(s)$ by dividing by $\varepsilon_i(s)$ all the elements of its

ith row ($i = 1, \ldots, \bar{q}$) so that

$$W_0(s) := \mathrm{diag}\left\{\frac{\varepsilon_1(s)}{\zeta(s)}, \ldots, \frac{\varepsilon_{\bar{q}}(s)}{\zeta(s)}\right\}\bar{Z}(s)\bar{T}^{-1}(s), \tag{3.3}$$

where $\bar{Z}(s)$ and $\bar{T}(s)$ are right coprime by the right coprimeness of $Z(s)$ and $T(s)$, $\bar{Z}(s)$ has no loss of rank at α_i and α_i^* ($i = 1, \ldots, \bar{\mu}$) and $\det \bar{T}(s)$ is a Hurwitz polynomial.

If $\det \bar{Z}(s)$ is not a Hurwitz polynomial, denote with $\gamma_i(s)$ the monic least common denominator of the poles of the ith column of $\bar{Z}^{-1}(s)$ which lie in the closed right half-plane ($i = 1, \ldots, \bar{q}$). Denote with $\varphi(s)$ the monic polynomial having α_i and α_i^*, $i = 1, \ldots, \bar{\mu}$, as its only roots, all simple ones; and consider the rational matrix

$$\bar{T}(s)\bar{Z}^{-1}(s)\,\mathrm{diag}\left\{\frac{\gamma_1(s)}{\varphi(s)}, \ldots, \frac{\gamma_{\bar{q}}(s)}{\varphi(s)}\right\}.$$

If such a matrix is not strictly proper, choose $\bar{q}$ monic Hurwitz polynomials $\varphi_1(s), \ldots, \varphi_{\bar{q}}(s)$ of suitable degrees so that the rational matrix

$$G_{MD}(s) := \bar{T}(s)\bar{Z}^{-1}(s)\,\mathrm{diag}\left\{\frac{\gamma_1(s)}{\varphi(s)\varphi_1(s)}, \ldots, \frac{\gamma_{\bar{q}}(s)}{\varphi(s)\varphi_{\bar{q}}(s)}\right\}, \tag{3.4}$$

is strictly proper. Let K_{MD} be a reachable and observable compensator having $G_{MD}(s)$ as transfer matrix and $r_{MD}(t) \in \mathbb{R}^{\bar{q}}$ as input. By construction, the only unstable poles of $G_{MD}(s)$ are the roots of $\varphi(s)$, which are not transmission zeros of $W_0(s)$. Hence, by a straightforward extension of Corollary 9.4 in Chen (1984) to detectability and stabilizability, the series connection Σ_D of K_{MD} followed by Σ_0 is stabilizable and detectable from $\bar{y}(t)$ at $\beta = \beta_0$.

By (3.3) and (3.4), the transfer matrix $G_D(s)$ of Σ_D from $r_{MD}(t)$ to $\bar{y}(t)$ at $\beta = \beta_0$ is diagonal and nonsingular, and expressed by

$$G_D(s) = \mathrm{diag}\left\{\frac{\varepsilon_1(s)\gamma_1(s)}{\zeta(s)\varphi(s)\varphi_1(s)}, \ldots, \frac{\varepsilon_{\bar{q}}(s)\gamma_{\bar{q}}(s)}{\zeta(s)\varphi(s)\varphi_{\bar{q}}(s)}\right\}. \tag{3.5}$$

Therefore, consider the control system depicted in Fig. 3, and choose a compensator K_{sD} with a diagonal transfer matrix $G_{sD}(s)$, such that the overall control system Σ is asymptotically stable at $\beta = \beta_0$. The existence of such a K_{sD} is guaranteed by the detectability and stabilizability of Σ_D at $\beta = \beta_0$ and by the diagonality of $G_D(s)$ at $\beta = \beta_0$. Notice that the $\bar{q}$ numerators $\chi_1(s), \ldots, \chi_{\bar{q}}(s)$ of the diagonal elements of $G_{sD}(s)$ must be nonzero at α_i and α_i^* ($i = 1, \ldots, \bar{\mu}$), otherwise a cancellation with $\varphi(s)$ should prevent Σ to be asymptotically stable at $\beta = \beta_0$. With this choice of K_{sD}, requirements (a) and (d) are met. By the same reason as in the proof of Theorem 1, (a) is preserved in some neighbourhood $\bar{\Psi}_c \subset \Omega$ of β_0.

Since, for each $i = 1, \ldots, \bar{q}$, $\gamma_i(s)$ and $\varphi(s)$ are coprime, $\chi_i(s)$ and $\varphi(s)$ are coprime, and the α_js ($j = 1, \ldots, \bar{\mu}$) are neither roots of $\det \bar{T}(s)$ nor of $\det \bar{Z}(s)$, it is readily seen that $\varphi(s)$ divides all $\bar{q}$ denominator polynomials in the Smith–McMillan form of $G(s) := G_{MD}(s)G_{sD}(s)$ (e.g. use the valuations of $G(s)$ at α_j and α_j^*, $j = 1, \ldots, \bar{\mu}$ (Kailath, 1980)). By well-known results (e.g. Francis and Wonham, 1975a; Grasselli and Nicolò, 1971) this proves that $\bar{e}(t) := r(t) - \bar{y}(t)$ asymptotically goes to zero for each reference signal $r(\cdot) \in \mathcal{R}$, for each $d_i(\cdot) \in \mathcal{D}_i$ ($i = 1, \ldots, \bar{\mu}$) and for $d_i(\cdot) = 0$, $i = \bar{\mu} + 1, \ldots, \mu$, at $\beta = \beta_0$ and for all $\beta \in \bar{\Psi}_c$. In addition, for each $\beta \in \bar{\Psi}_c$, for each reference signal $r(\cdot) \in \mathcal{R}$, for each $d_i(\cdot) \in \mathcal{D}_i$ ($i = 1, \ldots, \bar{\mu}$), and for $d_i(\cdot) = 0$, $i = \bar{\mu} + 1, \ldots, \mu$, the response $r_0(\cdot)$ is the sum of a signal $r_{0a}(\cdot) \in \mathcal{R}$ and a free response $r_{0b}(\cdot)$ of Σ corresponding to the same value of β. Therefore, the property of Σ_0 to satisfy requirements (a), (b), and (c) in some neighbourhood Ψ_0 of β_0, implies that for each $r(\cdot) \in \mathcal{R}$, for each $d_i(\cdot) \in \mathcal{D}_i$ ($i = 1, \ldots, \bar{\mu}$), for $d_i(\cdot) = 0$ ($i = \bar{\mu} + 1, \ldots, \mu$) and for each $\beta \in \Psi_0 \cap \bar{\Psi}_c$ the response $e_0(t)$ asymptotically goes to zero, as well as the response $\bar{e}(t)$ does. Hence $e(t)$ asymptotically goes to zero for each $r(\cdot) \in \mathcal{R}$, for each $d_i(\cdot) \in \mathcal{D}_i$ ($i = 1, \ldots, \bar{\mu}$), for $d_i(\cdot) = 0$ ($i = \bar{\mu} + 1, \ldots, \mu$) and for each $\beta \in \Psi_0 \cap \bar{\Psi}_c$.

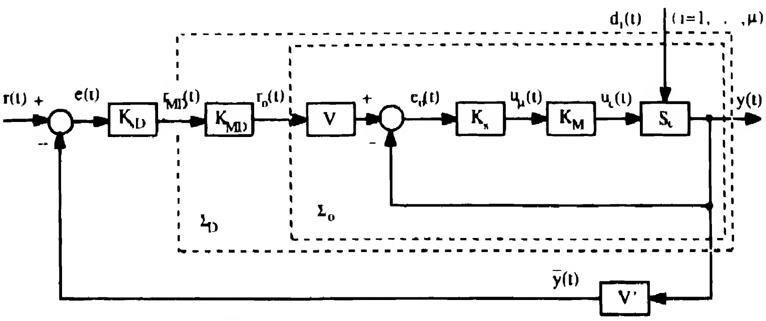


FIG. 3. The overall control system Σ for Problems 2, 3 and 4.

Lastly, notice that for each $\beta \in \bar{\Psi}_r$ the steady-state response of Σ to any $r(\cdot) \in \mathcal{R}$ and $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \mu)$ is unique. Therefore, the property of Σ_0 to satisfy requirements (a), (b) and (c) in Ψ_0 implies that for each $d_i(\cdot) \in \mathcal{D}_i$ $(i = \bar{\mu} + 1, \ldots, \mu)$ for $r(\cdot) = 0$ and for each $\beta \in \Psi_0 \cap \bar{\Psi}_c$, $r_0(t)$ and $e_0(t)$ asymptotically go to zero, whence $y(t)$ does.

This proves that requirement (b) is satisfied at $\beta = \beta_0$, and preserved, together with (a), in the neighbourhood $\Psi := \Psi_0 \cap \bar{\Psi}_c$ of β_0.

Remark 3.1 Requirements (a) and (b), as far as the only $\bar{y}(t)$ and the only reference signals $r(\cdot) \in \mathcal{R}$ and disturbance functions $d_i(\cdot) \in \mathcal{D}_i$ $(i = 1, \ldots, \bar{\mu})$ are concerned, are preserved for all $\beta \in \bar{\Psi}_c$, by virtue of the, say, "$\bar{q}$-dimensional" fully redundant internal model of such $d_i(\cdot)$ and $r(\cdot)$ contained in K_{MD} through $\varphi(s)$. Requirement (b) for $y_{q+1}(t), \ldots, y_q(t)$ and/or for disturbances $d_i(\cdot) \in \mathcal{D}_i$ $(i = \bar{\mu} + 1, \ldots, \mu)$ is preserved only for all $\beta \in \Psi_0 \cap \bar{\Psi}_c$. If, although $\bar{\mu} < \mu$, (2.7b) holds with μ instead of $\bar{\mu}$, then the same design procedure of K_M ensures $W_0(s)$ to have no transmission zeros at α_i and α_i^* for each $i = 1, \ldots, \mu$, thus allowing factors $(s - \alpha_i)$ and $(s - \alpha_i^*)$ to be contained in $\varphi(s)$ also for $i = \bar{\mu} + 1, \ldots, \mu$. In such a case, (b) is preserved for $\bar{y}(t)$ for all $\beta \in \bar{\Psi}_c$, and it is preserved for $y_{\bar{q}+1}(t), \ldots, y_q(t)$ for all $\beta \in \Psi_0 \cap \bar{\Psi}_c$.

The approach here followed in the design of $G_{MD}(s)$ for decoupling Σ_D at $\beta = \beta_0$ is the same as in Chen (1984). Notice also that the design of the diagonal elements of $G_{sD}(s)$ can be performed with the help of classical SISO techniques (Bode plots, etc.) in order to satisfy for Σ also the additional requirement (e), thus providing a solution not only of Problem 2, but also of Problem 3. Notice also that the role of $\zeta(s)$ and $\varepsilon_i(s)$ $(i = 1, \ldots, \bar{q})$ and the use of the choice of $\varphi_i(s)$ $(i = 1, \ldots, \bar{q})$ can be that of reducing the control effort of "cancelling" in part Σ_0.

As regards Problem 4, assume that, in the design of Σ_0, Ψ_r has been widened to some "large" closed (and possibly, prescribed) subset Ψ_s of Ω through the use for K_r of some robust stabilization design procedure, as mentioned in Remark 2.1. The transfer matrix $G_D(\beta, s)$ of Σ_D from $r_{MD}(t)$ to $\bar{y}(t)$ is square and nonsingular at $\beta = \beta_0$ (see $G_D(s)$ in (3.5)); since $G_D(\beta, s)$ is strictly proper (by the choice of $\varphi_1(s), \ldots, \varphi_{\bar{q}}(s)$) and has the same number of unstable modes for all $\beta \in \Psi_s$, then for all $\beta \in \Psi_s$ the perturbations $G_D(\beta, s)$ of the nominal $G_D(s)$ satisfy the assumptions in Doyle and Stein (1981). They can actually be expressed in the form of "unstructured" uncertainties at the output $\bar{y}(t)$, namely

$$G_D(\beta, s) = [I + \Delta(\beta, s)]G_D(s), \tag{3.6}$$

for some $\Delta(\beta, s)$ which can be computed. If the elements of $A(\beta)$, $B(\beta)$, $\hat{C}(\beta)$ and $\hat{D}(\beta)$ are continuous functions of β in Ψ_s, then $\Delta(\beta, s)$ satisfies a bound of the form

$$\bar{\sigma}[\Delta(\beta, j\omega)] < l_m(\omega), \quad \forall \beta \in \Psi_s, \quad \forall \omega \geq 0, \tag{3.7}$$

for some $l_m(\omega)$ which can be computed, where $\bar{\sigma}$ means maximum singular value. Therefore, the design of the diagonal elements of $G_{sD}(s)$ can be performed on the basis of the diagonal $G_D(s)$ expressed by (3.5) and of $l_m(\omega)$, with the help of SISO techniques, not only for stabilizing Σ at $\beta = \beta_0$ and for satisfying requirement (e), but also in order to satisfy the following conditions (Doyle and Stein, 1981), whose fulfilment is made easy by the diagonality of $G_D(s)$ and $G_{sD}(s)$:

$$\bar{\sigma}[G_D(j\omega)G_{sD}(j\omega)(I + G_D(j\omega)G_{sD}(j\omega))^{-1}] < 1/l_m(\omega) \quad \forall \omega \in [0, \infty], \tag{3.8}$$

$$\underline{\sigma}[G_D(j\omega)G_{sD}(j\omega)] \geq ps(\omega)/(1 - l_m(\omega)) \quad \forall \omega : l_m(\omega) < 1, \quad \underline{\sigma}[G_D(j\omega)G_{sD}(j\omega)] \gg 1, \tag{3.9}$$

(where $\underline{\sigma}$ means minimum singular value) for some $ps(\omega)$ characterizing a performance requirement on $\bar{y}(t)$ of the form

$$\bar{\sigma}[(I + G_D(\beta, j\omega)G_{sD}(j\omega))^{-1}] \leq 1/ps(\omega) \quad \forall \omega \in [0, \omega_0], \quad \text{for some } \omega_0 > 0. \tag{3.10}$$

Thus, relation (3.8) will guarantee the asymptotic stability of Σ for all $\beta \in \Psi_s$ (i.e. $\bar{\Psi}_c = \Psi_s$), and relation (3.9) will guarantee the fulfillment of the performance requirement (3.10) for all $\beta \in \Psi_s$. In this way all requirements (a), (b), (c), (d), (e) and (f) are met; in particular, requirement (b) remains valid in $\Psi_0 \cap \Psi_s = \Psi_0$, while, for $d_i(\cdot) = 0$ $(i = \bar{\mu} + 1, \ldots, \mu)$, as far as $\bar{y}(t)$ is concerned, it is preserved in all Ψ_s.

The main loop in Fig. 3 could be called the tracking loop, while the inner loop could be called the regulation loop. The order of K can be reduced if (2.26) holds.

Remark 3.2. Taking into account the discussion of Problem 4 contained in the previous remark, together with the sufficiency proof of Theorem 2, if $\bar{q} > 0$ and Assumption 1 and conditions (i) and (ii) hold, a solution K of Problem 4 can be obtained through the following design procedure.

Steps 1 and 2. As in Remark 2.2.

Step 3. Find a compensator K_s so that the system Σ_0 in Fig. 3 is asymptotically stable in some "large" and closed subset Ψ_s of Ω containing β_0 as an interior point (e.g. Doyle and Stein, 1981; Bernstein and Haddad, 1990).

Step 4. Find an irreducible right polynomial factorization $Z(s)T^{-1}(s)$ of the transfer matrix $W_0(s)$ of Σ_0 from $r_0(t)$ to $\bar{y}(t)$ at $\beta = \beta_0$ (Chen, 1984; Kailath, 1980).

Step 5. Compute the g.c.d. $\zeta(s)$ of all elements of $T(s)$, and, for each $i = 1, \ldots, \bar{q}$, the g.c.d. $\varepsilon_i(s)$ of all the elements of the ith row of $Z(s)$. Define $\bar{T}(s) := T(s)/\zeta(s)$, and $\bar{Z}(s)$ through (3.3).

Step 6. For each $j = 1, \ldots, \bar{q}$, compute the least common denominator $\gamma_j(s)$ of the poles of the jth column of $\bar{Z}^{-1}(s)$ which lie in the closed right half-plane. Compute the monic polynomial $\varphi(s)$ having α_i and α_i^* ($i = 1, \ldots, \mu$ if (2.7b) holds with μ instead of $\bar{\mu}$; $i = 1, \ldots, \bar{\mu}$ otherwise) as its only roots, all simple ones. Choose $\bar{q}$ Hurwitz polynomials $\varphi_1(s), \ldots, \varphi_{\bar{q}}(s)$ of minimal degrees so that matrix $G_{MD}(s)$ in (3.4) is strictly proper. Find a minimal realization K_{MD} of $G_{MD}(s)$.

Step 7. Compute the transfer matrix $G_D(\beta, s)$ of Σ_D from $r_{MD}(t)$ to $\bar{y}(t)$, and the bound $l_m(\omega)$ in (3.7) for $\Delta(\beta, j\omega)$ with the help of (3.5) and (3.6).

Step 8. Find the diagonal elements of the diagonal matrix $G_{sD}(s)$ with the help of SISO techniques so that the single loop requirements in (e) are met together with (a) and conditions (3.8) and (3.9) (rewritten in terms of the diagonal elements of $G_D(j\omega)$ and $G_{sD}(j\omega)$).

Step 5 only has the purpose of reducing the cancellations needed between $W_0(s)$ and $G_{MD}(s)$. It can be omitted and substituted with the following one.

Step 5'. Put $\bar{Z}(s) = Z(s)$ and $\bar{T}(s) = T(s)$.

4. CONCLUSIONS

The method here presented allows us to obtain decoupling at the nominal parameters and a robust asymptotic tracking and disturbance rejection and to robustly satisfy a performance requirement in the case when it is known the actual dependence of the plant to be controlled on some "physical" parameters, and the scalar outputs which must track corresponding reference signals are a proper subset of the set of the controlled outputs (in which case no solution may exist with the existing methods).

REFERENCES

Ackermann, J. (1980). Parameter space design of robust control systems. *IEEE Trans. Aut. Control*, **AC-25**, 1058–1072.

Barmish, B. R. (1985). Necessary and sufficient conditions for quadratic stabilizability of an uncertain system. *J.O.T.A.*, **46**, 399–408.

Basile, G. and G. Marro (1991). *Controlled and Conditioned Invariants in Linear System Theory*. Prentice-Hall, Englewood Cliffs, NJ.

Basile, G., G. Marro and A. Piazzi (1987). Revisiting the regulator problem in the geometric approach. Part II—Asymptotic tracking and regulation in the presence of disturbances. *J. of Opt. Theory and Appl.*, **53**, 1.

Bernstein, D. S. and W. M. Haddad (1990). Robust stabilization and performance via fixed-order dynamic compensation with guaranteed cost bounds. *Mathematics of Control, Signals, and Systems*, 139–163.

Chen, C. T. (1984). *Linear System Theory and Design*. Holt, Rinehart and Winston, New York.

Colaneri, P. (1990). Discrete-time linear periodic systems: stabilization and zero error regulation. *Preprints of the 11th IFAC World Congress*, Tallinn, Estonia (SU), Vol. 2, pp. 144–149.

Davison, E. J. (1976). The robust control of a servomechanism problem for linear time-invariant multivariable systems. *IEEE Trans. Aut. Control*, **AC-21**, 25–34.

Davison, E. J. and A. Goldenberg (1975). Robust control of a general servomechanism problem: the servo compensator. *Automatica*, **11**, 461–471.

Davison, E. J. and S. H. Wang (1974). Properties and calculation of transmission zeros of linear multivariable systems. *Automatica*, **10**, 643–658.

Desoer, C. A. and Y. T. Wang (1978). On the minimum order of a robust servocompensator. *IEEE Trans. Aut. Control*, **23**, 70–73.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. Aut. Control*, **AC-26**, 4–16.

Dorato, P. (1987). *Robust Control*. IEEE Press, New York.

Francis, B. A. (1977). The linear multivariable regulator problem. *SIAM J. Control*, **15**, 486–505.

Francis, B. A. (1987). *A Course in H_∞ Control Theory*. Springer-Verlag, Berlin.

Francis, B. A. and W. M. Wonham (1975a). The role of transmission zeros in linear multivariable regulators. *Int. J. Control*, **22**, 657–681.

Francis, B. A. and W. M. Wonham (1975b). The internal model principle for linear multivariable regulators. *Appl. Math. Opt.*, **2**, 170–194.

Francis, B. A. and W. M. Wonham (1976). The internal model principle of control theory. *Automatica*, **12**, 457–465.

Galimidi, A. R. and B. R. Barmish (1986). The constrained Lyapunov problem and its application to robust output feedback stabilization. *IEEE Trans. Aut. Control*, **AC-31**, 410–419.

Grasselli, O. M., A. Isidori and F. Nicolò (1979). Output regulation of a class of bilinear systems under constant disturbances. *Automatica*, **15**, 189–195.

Grasselli, O. M. and S. Longhi (1991a). Robust output regulation under uncertainties of physical parameters. *Systems & Control Letters*, **16**, 33–40.

Grasselli, O. M. and S. Longhi (1991b). Robust tracking and regulation of linear periodic discrete-time systems. *Int. J. Control*, **54**, 613–633.

Grasselli, O. M. and F. Nicolò (1971). Modal synthesis of astatic multivariable regulation systems. *Proc. 2nd IFAC*

Symp. on Multiv. Tech. Control Syst., Dusseldorf, paper 1.1.4.

Grasselli, O. M. and F. Nicolò (1973a). Multivariable control systems with a structural steady-state signal invariance. *Proc. 3rd IFAC Symp. on Sensitivity, Adaptivity and Optimality*, Ischia (Italy), pp. 85–90.

Grasselli, O. M and F. Nicolò (1973b). Steady-state invariant control systems under polynomial disturbances. *Ricerche di Automatica*, **4**, 105–141.

Grasselli, O. M and F. Nicolò (1976). Steady-state invariant control systems under disturbances satisfying differential equations *J. Franklin Inst.*, **301**, 287–305.

Gu, K., Y H. Chen, M. A. Zohdy and N. K. Loh (1991). Quadratic stabilizability of uncertain systems: a two-level optimization setup. *Automatica*, **27**, 161–165.

Kailath, T. (1980). *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Keel, L. H., S. P. Bhattacharyya and Jo W. Howze (1988). Robust control with structured perturbations. *IEEE Trans. Aut. Control*, **AC-33**, 68–77.

Khargonekar, P. P., I. R. Petersen and K. Zhou (1990). Robust stabilization of uncertain linear systems: quadratic stabilizability and H^∞ control theory. *IEEE Trans. Aut. Control*, **35**, 356–361.

Petersen, I R. and C V. Hollot (1986) A Riccati approach to the stabilization of uncertain linear systems. *Automatica*, **22**, 397–411.

Safonov, M. G., A. J. Laub and G. L. Hartmann (1981). Feedback properties of multivariable systems: the role and use of the return difference matrix *IEEE Trans Aut. Control*, **AC-26**, 47–65.

Siljak, D. D. (1989). Parameter space methods for robust control design: a guided tour. *IEEE Trans. Aut. Control*, **34**, 674-688.

Staats, P W. Jr and J P. Pearson (1977). Robust solution of the linear servomechanism problem. *Automatica*, **13**, 125–138.

Vidyasagar, M. and H. Kimura (1986) Robust controllers for uncertain linear multivariable systems *Automatica*, **22**, 85–94

Wei, K and R K. Yedavalli (1989) Robust stabilization for linear systems with both parameter variation and unstructured uncertainty. *IEEE Trans. Aut. Control*, **AC-34**, 149–156.

Willems, J. L. and J. C. Willems (1983). Robust stabilization of uncertain systems. *SIAM J. Control Optim.*, **21**, 352–374.

Wonham, W. M. (1974). *Linear Multivariable Control—A Geometric Approach*. Springer-Verlag, Berlin.

Young, P. C. and J. C. Willems (1972). An approach to the linear multivariable servomechanism problem. *Int. J. Control*, **15**, 961–979.

Zhou, K. and P. P. Khargonekar (1988). Robust stabilization of linear systems with norm-bounded time-varying uncertainty. *Syst. and Control Letters*, **10**, 17–20.

APPENDIX. PROOF OF LEMMA 1

Define

$$S(s,\beta) := \begin{bmatrix} A(\beta) - sI & B(\beta) \\ C(\beta) & D(\beta) \end{bmatrix}, \tag{A.1}$$

$$\Gamma(\beta) := \operatorname{diag}\{S(\alpha_1,\beta), S(\alpha_2,\beta), \dots, S(\alpha_\mu,\beta)\}. \tag{A.2}$$

The lemma will be proved by showing that the hypothesis implies the existence of an interior point β_0 of $\bar{\Psi}_a$ and a neighbourhood Ψ_a of β_0, $\Psi_a \subset \bar{\Psi}_a$, such that

$$\operatorname{rank}\Gamma(\beta) = \operatorname{rank}\Gamma(\beta_0) \quad \forall \beta \in \Psi_a. \tag{A.3}$$

Now, by the continuity of $A(\beta)$, $B(\beta)$, $C(\beta)$ and $D(\beta)$ in $\bar{\Psi}_a$, if (A.3) is violated for $\beta_0 = \bar{\beta}_0$ and for all neighbourhoods Ψ_a of $\bar{\beta}_0$, $\Psi_a \subset \bar{\Psi}_a$, then there exists β_1 arbitrarily near $\bar{\beta}_0$ such that

$$\operatorname{rank}\Gamma(\beta_1) > \operatorname{rank}\Gamma(\bar{\beta}_0), \tag{A.4}$$

and for which there exists a neighbourhood $\bar{\Psi}_1$ of β_1, $\bar{\Psi}_1 \subseteq \bar{\Psi}_a$, in which the elements of $A(\beta)$, $B(\beta)$, $C(\beta)$, $D(\beta)$, $\bar{C}(\beta)$ and $\bar{D}(\beta)$ are continuous.

The proof is completed by induction, by noting that the rank of $\Gamma(\beta)$ is bounded for all $\beta \subset \bar{\Psi}_a$, since $\Gamma(\beta)$ has finite dimensions.

Automatica, Vol 29, No 1, pp 181–189, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Robust Stabilization of Feedback Linearizable Time-varying Uncertain Nonlinear Systems*†

RICCARDO MARINO‡ and PATRIZIO TOMEI‡

A global robust stabilizing state feedback controller is provided for a class of single-input nonlinear systems with bounded unmodelled time-varying disturbances entering nonlinearly in the state equations.

Key Words—Robust stabilization, nonlinear systems, feedback linearization, self-tuning stabilization

Abstract—We consider single-input nonlinear systems with unknown unmodelled time-varying parameters or disturbances whose bounds are known Assuming that the undisturbed system is globally feedback linearizable and that a triangularity condition holds for the uncertain terms we design a robust global stabilizing state feedback control Disturbances are not required to enter linearly in the state equations When they do enter linearly, the stabilization problem can be solved without knowing bounds on disturbances by using a self-tuning version of the robust control In particular, any linear system in controller canonical form perturbed by unknown global Lipschitz nonlinearities satisfying triangularity conditions is shown to be globally stabilizable by a fixed dynamic state feedback compensator whose order equals the state space dimensions

1 INTRODUCTION

WE CONSIDER single-input nonlinear systems

$$\begin{aligned}\dot{x} &= f(x) + q(x, \theta(t)) + g(x)u \\ &= \bar{f}(x, \theta(t)) + g(x)u, \qquad (1)\\ &\quad x \in R^n,\ u \in R,\ \theta \in \Omega \subset R^p,\end{aligned}$$

where $\theta(t)$ is a vector of unknown, time-varying, piecewise continuous parameters or disturbances which takes values in the compact set $\Omega \subset R^p$, $g(x)$ is a smooth vector field with $g(x) \neq 0$, $\forall x \in R^n$, $f(x)$ is the nominal or undisturbed smooth vector field

$$f(x) = \bar{f}(x, \theta_N), \tag{2}$$

with θ_N the nominal constant value of θ. The function

$$q(x, \theta(t)) = \bar{f}(x, \theta(t)) - \bar{f}(x, \theta_N), \tag{3}$$

is assumed to be smooth and contains all the uncertainties in the system, including disturbances and uncertain nonlinear terms. We assume that there exists an isolated equilibrium point (taken without loss of generality to be the origin) which is not affected by the vector $\theta(t)$, i.e. $f(0) = 0$ and $q(0, \theta) = 0, \forall \theta \in \Omega$. We address the local (global) robust stabilization problem, i.e. the design of a smooth state feedback control which makes the origin locally (globally) asymptotically stable for every $\theta(t)$.

When $\theta(t)$ is a constant and known parameter vector, the stabilization problem, which is a fundamental issue in nonlinear control theory, has been extensively studied in the literature (see Sontag (1990) and Bacciotti (1992) for an extensive bibliography and a recent monograph). The class of locally (globally) feedback linearizable systems, i e those systems

$$\dot{x} = f(x) + g(x)u, \quad f(0) = 0, \quad x \in R^n, \tag{4}$$

which are transformable by a local (global) diffeomorphism

$$z = T(x), \quad T(0) = 0, \quad z \in \mathbb{R}^n, \tag{5}$$

and by a nonsingular state feedback

$$u = \alpha(x) + \beta(x)v, \quad \alpha(0) = 0, \tag{6}$$

with $\beta(0) \neq 0$ (for global feedback linearization $\beta(x) \neq 0, \forall x \in R^n$) into a linear controllable system

$$\dot{z} = Az + bv, \tag{7}$$

with

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \tag{8}$$

* Received 5 August 1991, received in final form 6 August 1992 The preliminary version of this paper was presented at the IFAC Symposium on Adaptive Systems in Control and Signal Processing which was held in Grenoble, France during 1–3 July 1992 The Published Proceedings of this IFAC Meeting may be ordered from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, U K This paper was recommended for publication in revised form by Guest Editor R F Curtain

† This work was supported by Ministero dell'Università e della Ricerca Scientifica e Tecnologica

‡ Dipartimento di Ingegneria Elettronica, Seconda Università di Roma, "Tor Vergata", Via O Raimondo, 00173 Roma, Italy

is locally (globally) stabilizable. Necessary and sufficient conditions for local feedback linearlization have been obtained in Jakubczyk and Respondek (1980) and Su (1982), while the problem of global feedback linearization is addressed in Boothby (1984, 1986), Dayawansa *et al.* (1985) and Respondek (1986).

Since feedback linearization techniques involve nonlinearity cancellation, the problem of extending those techniques to uncertain systems containing a vector of constant unknown parameters θ was addressed in Nam and Arapostathis (1988), Taylor *et al.* (1989), Sastry and Isidori (1989) and Kanellakopoulos *et al.* (1991a, b) where adaptive versions of feedback linearizing techniques were proposed under the linear parametrization assumption which restricts the vector θ to enter linearly in equations (1), namely

$$q(x, \theta) = \sum_{i=1}^{p} \theta_i q_i(x).$$

In Taylor *et al.* (1989) and Kanellakopoulos *et al.* (1991a, b) the nominal system is assumed to be globally feedback linearizable and the vector fields $q_i(x)$ are required to satisfy structural conditions: matching conditions were introduced in Taylor *et al.* (1989) and weakened by extended matching conditions in Kanellakopoulos *et al.* (1991a) and, recently, in Kanellakopoulos *et al.* (1991b) by the strict triangularity conditions which are the least restrictive ones. Those conditions do not impose any growth requirement on the nonlinearities, such as boundedness, sector or Lipschitz assumptions (see Nam and Arapostathis, 1988; Sastry and Isidori, 1989).

The main result of this paper, given in Section 3, leads to the design of a local (global) robust stabilizing state feedback control under three assumptions: the nominal system (f, g) is locally (globally) feedback linearizable; the uncertain vector $q(x, \theta(t))$ satisfies coordinate-free triangularity conditions; Ω is a known compact set. This extends the stabilization results presented in Kanellakopoulos *et al.* (1991b) since the crucial assumption of linear parametrization is removed and bounded time-varying parameters and uncertainties on non-linearities (such as look up tables) are allowed. The technique of proof is based on a recursive, constructive algorithm which simultaneously builds a Lyapunov function and a fixed static nonlinear state feedback control that makes the origin locally (globally) asymptotically stable for any $\theta(t)$. When the components of $q(x, \theta(t))$ satisfy, in suitable coordinates, more restrictive global Lipschitz conditions then the recursive algorithm provides a globally stabilizing linear controller in those coordinates, thus recovering earlier high-gain robust control schemes (Thorp and Barmish, 1981; Barmish *et al.*, 1983).

When the linear parametrization assumption is met, the robust stabilizing control can be made adaptive by the state feedback self-tuning algorithm given in Section 4 consisting of a dynamic compensator of order at most n which tunes the control parameters. This self-tuning algorithm allows us to remove the hypothesis that the compact set Ω is known. When global Lipschitz conditions are satisfied, Lipschitz constants need not to be known. In Section 4 it is shown that any linear system in controller canonical form perturbed by global Lipschitz nonlinearities with unknown Lipschitz constants can be globally stabilized by a fixed state feedback dynamic compensator of order n.

The reader is referred to Isidori (1989) and Nijmeijer and van der Schaft (1990) for basic notations and results in geometric nonlinear control theory.

2 PRELIMINARIES AND BASIC RESULTS

We first recall the necessary and sufficient conditions which identify those single-input systems which are locally (globally) feedback linearizable.

Theorem 2.1 (Jakubczyk and Respondek, 1980; Su, 1982). The system (4) is locally feedback linearizable, i.e. there exist a local diffeomorphism (5) and a nonsingular state feedback (6) such that the closed loop system (4)–(6) in z-coordinates becomes (7) if, and only if,

- $G^i = \text{span}\,\{g, \ldots, ad_f^i g\}$ is involutive of constant rank $i+1$ for $0 \le i \le n-1$.

A global version of Theorem 2.1 is given in Boothby (1986) (see also Boothby, 1984; Dayawansa *et al.*, 1985; Respondek, 1986).

Theorem 2.2 (Boothby, 1986). The system (4) is globally feedback linearizable, i.e. there exist a global diffeomorphism (5) and a nonsingular state feedback (6) such that the closed loop system (4)–(6) in z-coordinates becomes (7) if, and only if:

- it is locally feedback linearizable;
- there exists a function $h \in C^\infty(R^n)$ such that $dh \neq 0$, $\forall x \in R^n$, and $\langle dh, G^{n-2} \rangle = 0$, $\forall x \in R^n$;
- the vector fields $ad_{\tilde f}^i \tilde g$, $0 \le i \le n-1$, with $\tilde f = f - (L_g L_f^{n-1} h)^{-1} L_f^n h$, $\quad \tilde g = (L_g L_f^{n-1} h)^{-1} g$ are complete.

In what follows we shall always assume that the nominal system (f, g) is locally (globally) feedback linearizable. As far as the uncertainties are concerned we shall make the following structural coordinate-free assumption.

Strict triangularity assumption.

- $ad_q G^i \subset G^i, \quad 0 \le i \le n-2, \quad \forall \theta \in \Omega.$

Remark 2.1. The strict triangularity assumption was introduced in Kanellakopoulos *et al.* (1991b) (where it is called strict feedback) generalizing earlier structural conditions (called pure feedback in Kanellakopoulos *et al.*, 1991b) which were obtained in Akhrif and Blankenship (1988). In Kanellakopoulos *et al.* (1991b) and Akhrif and Blankenship (1988) the parameter vector is assumed to be constant and to enter linearly.

Lemma 2.1. If the nominal system (f, g) is locally (globally) feedback linearizable and the strict triangularity assumption is satisfied, then there exist a local (global) diffeomorphism (5) and a nonsingular state feedback (6) such that the closed loop system (1)–(6) in z-coordinates becomes

$$\begin{aligned} \dot z_j &= z_{j+1} + \phi_j(z_1, \dots, z_j, \theta(t)), \quad 1 \le j \le n-1, \\ \dot z_n &= v + \phi_n(z_1, \dots, z_n, \theta(t)). \end{aligned} \tag{9}$$

Proof. By assumption the distribution G^{n-2} is involutive and of constant rank $n-1$. By Frobenius Theorem there exists a function h such that in U_0, a neighborhood of the origin, (in R^n if Theorem 2.2 applies)

$$\langle dh, ad_f^{n-1} g\rangle \neq 0,$$
$$\langle dh, G^{n-2}\rangle = 0.$$

It follows that $dh, \dots, d(L_f^{n-1}h)$ are linearly independent and that

$$z_1 = h(x)$$
$$z_i = L_f^{i-1}h(x), \quad 2 \le i \le n,$$

is a local diffeomorphism in U_0 (or a global diffeomorphism if Theorem 2.2 applies) and $(z_1, \dots, z_n)$ are local (global) coordinates. In new coordinates system (1) becomes

$$\begin{aligned} \dot z_1 &= z_2 + L_q h, \\ &\vdots \\ \dot z_n &= L_f^n h + (L_g L_f^{n-1} h) u + L_q L_f^{n-1} h, \end{aligned}$$

and

$$G^0 = \operatorname{span}\left\{\frac{\partial}{\partial z_n}\right\},$$

$$G^i = \operatorname{span}\left\{\frac{\partial}{\partial z_n}, \dots, \frac{\partial}{\partial z_{n-i}}\right\},$$

$$G^{n-1} = \operatorname{span}\left\{\frac{\partial}{\partial z_n}, \dots, \frac{\partial}{\partial z_1}\right\}.$$

Since $ad_q G^i \subset G^i$, $0 \le i \le n-2$, it follows that

$$\left[L_q h \frac{\partial}{\partial z_1} + \dots + L_q L_f^{n-1} h \frac{\partial}{\partial z_n}, G^i\right] \subset G^i,$$

i.e.

$$L_q h = \phi_1(z_1, \theta(t)),$$
$$L_q L_f^i h = \phi_{i+1}(z_1, \dots, z_{i+1}, \theta(t)), \quad 1 \le i \le n-1.$$

In conclusion, setting $L_f^n h + (L_g L_f^{n-1} h) u = v$ we have (9).

3 ROBUST STABILIZATION

We now state and prove the main theorem which provides the design of robust state feedback stabilizing controls.

Theorem 3.1. Consider the system (1). If Ω is a known compact set, the nominal system (f, g) is locally (globally) feedback linearizable and the strict triangularity assumption is satisfied, then there exists a local (global) static state feedback stabilizing controller.

Proof. Since the assumptions of Lemma 2.1 are satisfied, system (1) can be locally (globally) transformed into

$$\begin{aligned} \dot z_1 &= z_2 + \phi_1(z_1, \theta(t)), \\ \dot z_2 &= z_3 + \phi_2(z_1, z_2, \theta(t)), \\ &\vdots \\ \dot z_n &= \phi_n(z_1, \dots, z_n, \theta(t)) + v. \end{aligned} \tag{10}$$

The proof proceeds by induction. We first prove the following claim.

Claim. Assume that for a given index i, $1 \le i < n$, for the system

$$\begin{aligned} \dot z_1 &= z_2 + \phi_1(z_1, \theta(t)), \\ &\vdots \\ \dot z_i &= v_i + \phi_i(z_1, \dots, z_i, \theta(t)), \end{aligned} \tag{11}$$

there exist i smooth functions

$$z_2^*(z_1), z_3^*(z_1, z_2), \dots, z_{i+1}^*(z_1, \dots, z_i),$$
$$z_j^*(0, \dots, 0) = 0, \quad 2 \le j \le i+1, \tag{12}$$

such that for the closed loop system with the control $v_i = z_{i+1}^*$ in the new coordinates

$$\begin{aligned} \tilde z_1 &= z_1, \\ \tilde z_2 &= z_2 - z_2^*(z_1), \\ &\vdots \\ \tilde z_i &= z_i - z_i^*(z_1, \dots, z_{i-1}), \end{aligned} \tag{13}$$

the function

$$V_i = \tfrac{1}{2}\sum_{j=1}^{i} \bar{z}_j^2, \tag{14}$$

has a time derivative satisfying the inequality

$$\dot{V}_i \le -\sum_{j=1}^{i} (k_j - i + 1)\bar{z}_j^2, \tag{15}$$

with $k_j - i > 0$, $1 \le j \le i$. Then, for the system

$$\begin{aligned} \dot{z}_1 &= z_2 + \phi_1(z_1, \theta(t)), \\ &\vdots \\ \dot{z}_{i+1} &= v_{i+1} + \phi_{i+1}(z_1, \ldots, z_{i+1}, \theta(t)), \end{aligned} \tag{16}$$

there exists a smooth control

$$v_{i+1} = z_{i+2}^*(z_1, \ldots, z_{i+1}), \quad z_{i+2}^*(0, \ldots, 0) = 0, \tag{17}$$

such that for the closed loop system in new coordinates

$$\bar{z}_j, \quad 1 \le j \le i, \quad \bar{z}_{i+1} = z_{i+1} - z_{i+1}^*(z_1, \ldots, z_i), \tag{18}$$

the function

$$V_{i+1} = \tfrac{1}{2}\sum_{j=1}^{i+1} \bar{z}_j^2, \tag{19}$$

has time derivative satisfying the inequality

$$\dot{V}_{i+1} \le -\sum_{j=1}^{i+1} (k_j - i)\bar{z}_j^2, \tag{20}$$

with $k_{i+1} - i > 0$.

Proof of the claim. Consider the system (16). Make the change of coordinates given by (13) and (18). According to (15) and (19), the function V_{i+1} is such that

$$\dot{V}_{i+1} \le -\sum_{j=1}^{i} (k_j - i + 1)\bar{z}_j^2 + \bar{z}_i\bar{z}_{i+1} + \bar{z}_{i+1}\dot{\bar{z}}_{i+1}. \tag{21}$$

From (12) and (16), we have

$$\dot{\bar{z}}_{i+1} = z_{i+2} + \phi_{i+1}(z_1, \ldots, z_{i+1}, \theta(t)) - \sum_{j=1}^{i} \frac{\partial z_{i+1}^*}{\partial z_j}\dot{z}_j. \tag{22}$$

By virtue of (11), (13) and (18), we can write

$$\phi_{i+1}(z_1, \ldots, z_{i+1}, \theta(t)) - \sum_{j=1}^{i} \times \frac{\partial z_{i+1}^*}{\partial z_j}\dot{z}_j \triangleq \psi_{i+1}(\xi_{i+1}, \theta(t)), \tag{23}$$

where $\xi_i = [\bar{z}_1, \ldots, \bar{z}_i]^T$ and

$$\psi_{i+1}(0, \theta) = 0, \quad \forall \theta \in \Omega. \tag{24}$$

Since ψ_{i+1} is a smooth function of ξ_{i+1} and (24) holds, we can write (see Nijmeijer and van der Schaft, 1990)

$$\psi_{i+1}(\xi_{i+1}, \theta(t)) = \sum_{j=1}^{i+1} \bar{\psi}_j(\xi_{i+1}, \theta(t))\bar{z}_j, \tag{25}$$

with $\bar{\psi}_j$, $1 \le j \le i+1$, continuous functions. Since $\theta \in \Omega$, a known compact set, we can find a smooth function $\alpha_{i+1}(\xi_{i+1})$ such that

$$|\bar{\psi}_j(\xi_{i+1}, \theta(t))| \le \frac{\alpha_{i+1}(\xi_{i+1})}{i+1}, \quad 1 \le j \le i+1, \quad \forall \theta \in \Omega. \tag{26}$$

Define z_{i+2}^* as

$$z_{i+2}^* = -\bar{z}_i - (k_{i+1} - i + 1)\bar{z}_{i+1} - \bar{z}_{i+1}\alpha_{i+1}^2(\xi_{i+1}), \tag{27}$$

with $k_{i+1} - i > 0$. Taking (21)–(23), (26) and (27) into account, with $z_{i+2} = z_{i+2}^*$ in (16), we obtain

$$\begin{aligned} \dot{V}_{i+1} &\le -\sum_{j=1}^{i+1} (k_j - i + 1)\bar{z}_j^2 - \bar{z}_{i+1}^2\alpha_{i+1}^2 + |\bar{z}_{i+1}|\,\|\xi_{i+1}\|\,\alpha_{i+1} \\ &\le -\sum_{j=1}^{i+1} (k_j - i)\bar{z}_j^2 - \begin{bmatrix} \|\xi_{i+1}\| \\ |\bar{z}_{i+1}| \end{bmatrix}^T \times \begin{bmatrix} 1 & -\frac{\alpha_{i+1}}{2} \\ -\frac{\alpha_{i+1}}{2} & \alpha_{i+1}^2 \end{bmatrix} \begin{bmatrix} \|\xi_{i+1}\| \\ |\bar{z}_{i+1}| \end{bmatrix} \\ &\le -\sum_{j=1}^{i+1} (k_j - i)z_j^2, \end{aligned} \tag{28}$$

which concludes the proof of the claim.

We now show that the claim is true for $i = 1$. Consider the system

$$\dot{z}_1 = z_2 + \phi_1(z_1, \theta(t)). \tag{29}$$

Since ϕ_1 is a smooth function of z_1 and $\phi_1(0, \theta) = 0$, $\forall \theta \in \Omega$, we can write (see Nijmeijer and van der Schaft, 1990)

$$\phi_1(z_1, \theta(t)) = z_1\psi_1(z_1, \theta(t)), \tag{30}$$

where ψ_1 is a continuous function. Since $\theta(t) \in \Omega$, we can find a smooth function $\alpha_1(z_1)$ such that

$$\psi_1(z_1, \theta(t)) \le \alpha_1(z_1), \quad \forall \theta \in \Omega. \tag{31}$$

Let z_2^* be defined as

$$z_2^* = -k_1z_1 - z_1\alpha_1(z_1), \tag{32}$$

with $k_1 > 0$. Consider the function

$$V_1 = \tfrac{1}{2}z_1^2. \tag{33}$$

The time derivative of (33) computed with $z_2 = z_2^*$ in (29), is given by

$$\dot{V}_1 = -k_1 z_1^2 - z_1^2(\alpha_1(z_1) - \psi_1(z_1, \theta(t)), \quad (34)$$

and, taking (30), (31) into account, is such that

$$\dot{V}_1 \le -k_1 z_1^2. \quad (35)$$

Applying the claim $(n-1)$-times we can construct a smooth function $z_{n+1}^*(z_1, \ldots, z_n)$ which determines the control

$$v = v_n = z_{n+1}^*(z_1, \ldots, z_n), \quad (36)$$

and a global change of coordinates

$$\begin{aligned} \bar{z}_1 &= z_1, \\ \bar{z}_2 &= z_2 - z_2^*(z_1), \\ &\vdots \\ \bar{z}_n &= z_n - z_n^*(z_1, \ldots, z_{n-1}), \end{aligned} \quad (37)$$

such that for the closed loop system (10), (36) the radially unbounded, positive definite function

$$V_n = \tfrac{1}{2}\sum_{i=1}^{n} \bar{z}_i^2, \quad (38)$$

has time derivative satisfying the inequality

$$\dot{V}_n \le -c\,\|\bar{z}\|^2, \quad (39)$$

with $c > 0$. As a result, the equilibrium point $\bar{z} = 0$ is globally uniformly asymptotically stable (Hahn, 1967). Since $\bar{z}$ is related to x by the transformations (5) and (37) it follows that $x = 0$ is a locally (globally) uniformly asymptotically stable equilibrium point for system (1), (36).

Remark 3.1. As far as the stabilization problem is concerned, under the assumption that Ω is known, Theorem 3.1 generalizes a result obtained in Kanellakopoulos *et al.* (1991b). In fact, while the parameter vector is supposed to be constant and to enter linearly in Kanellakopoulos *et al.* (1991b), this is not required in Theorem 3.1. Moreover, as shown in (26) and (31), uncertainties in the nonlinear terms (e.g. look up tables) are allowed in Theorem 3.1 as long as functional bounds are known. On the other hand in Kanellakopoulos *et al.* (1991b) the tracking problem is solved as well while Theorem 3.1 only addresses the stabilization problem.

Remark 3.2. Assume that the origin is not an equilibrium point independent of $\theta(t)$, i.e. assume that

$$q(0, \theta(t)) \ne 0, \quad \text{for some } \theta \in \Omega.$$

Following the proof of Theorem 3.1, system (1) is transformed into (10) where now

$$\phi_j(0, \ldots, 0, \theta(t)) \ne 0, \quad \text{for some } \theta \in \Omega.$$

We rewrite system (10) as

$$\begin{aligned} \dot{z}_j &= z_{j+1} + [\phi_j(z_1, \ldots, z_j, \theta(t)) \\ &\quad - \phi_j(0, \ldots, 0, \theta(t))] + \phi_j(0, \ldots, 0, \theta(t)) \\ &\triangleq z_{j+1} + \bar{\phi}_j(z_1, \ldots, z_j, \theta(t)) \\ &\quad + \phi_j(0, \ldots, 0, \theta(t)), \quad 1 \le j \le n-1, \\ \dot{z}_n &= [\phi_n(z_1, \ldots, z_n, \theta(t)) - \phi_n(0, \ldots, 0, \theta(t))] \\ &\quad + \phi_n(0, \ldots, 0, \theta(t)) + v \\ &\triangleq \bar{\phi}_n(z_1, \ldots, z_n, \theta(t)) + \phi_n(0, \ldots, 0, \theta(t)) + v, \end{aligned}$$

and design the robust stabilizing controller considering only the terms $\bar{\phi}_j(z_1, \ldots, z_j, \theta(t))$ which vanish in the origin. The terms $\phi_j(0, \ldots, 0, \theta(t))$ which act on the system as additive bounded (since $\theta(t)$ is bounded and ϕ_j is smooth) disturbances are ignored. Since the final controlled system is exponentially stable in $\bar{z}$-coordinates (both V and $\dot{V}$ are quadratic) the additive disturbances which have been neglected in the controller design will produce bounded perturbations on the state in $\bar{z}$-coordinates and, therefore, on the state in z-coordinates when the diffeomorphism $z = T(x)$ is global.

Corollary 3.1. Consider the nonlinear system

$$\dot{z} = Az + bu + \begin{bmatrix} \phi_1(z, \theta(t)) \\ \vdots \\ \phi_n(z, \theta(t)) \end{bmatrix}, \quad z \in R^n, \quad (40)$$

with (A, b) in controller canonical form (8). If

$$|\phi_i(z, \theta(t))| \le \mu_i \, \|[z_1 \ldots, z_i]^T\|, \quad 1 \le i \le n, \quad (41)$$

with *known* Lipschitz constants μ_i, then there exists a *static linear* globally stabilizing state feedback controller.

Proof. It is a direct implication of the proof of Theorem 3.1 since the functions $\alpha_1, \ldots, \alpha_n$ which satisfy (26) and (31) can be chosen as the constants μ_i.

Example 3.1. Consider the system

$$\begin{aligned} \dot{x}_1 &= x_2 + x_1 \sin t^2, \\ \dot{x}_2 &= u. \end{aligned} \quad (42)$$

Since $|x_1 \sin t^2| \le |x_1|$ Corollary 3.1 applies with $\mu_1 = 1$, $\mu_2 = 0$ and therefore we can design a linear stabilizing controller. Following the proof of Theorem (3.1) (see (32)) we define

$$x_2^* = -k_1 x_1 - x_1.$$

The control u, according to (27) and (36), is

chosen as

$$u = -x_1 - k_2(x_2 - x_2^*)$$
$$\quad x_1 - k_2(k_1 + 1)x_1 - k_2 x_2. \qquad (43)$$

4. SELF-TUNING STABILIZATION

In this section we analyse the special case in which the parameter vector θ appears linearly.

Linear parametrization assumption.

- The vector field $q(x, \theta(t))$ is expressed as

$$q(x, \theta(t)) = \sum_{i=1}^{p} \theta_i(t) q_i(x).$$

If the linear parametrization assumption is satisfied we can construct a dynamic controller without requiring the compact set Ω to be known.

Theorem 4.1. Consider the system (1). Assume that the nominal system (f, g) is locally (globally) feedback linearizable and that the strict triangularity and the linear parametrization assumptions are satisfied; then there exists a local (global) dynamic state feedback stabilizing controller.

Proof. Lemma 2.1 applies and system (1) can be locally (globally) transformed into the form (10) which by virtue of the linear parametrization assumption becomes

$$\dot{z}_1 = z_2 + \sum_{i=1}^{p} \theta_i(t)\phi_{1,i}(z_1),$$
$$\vdots \qquad (43)$$
$$\dot{z}_n = v + \sum_{i=1}^{p} \theta_i(t)\phi_{n,i}(z_1, \ldots, z_n).$$

Now, the proof proceeds by induction.

Claim. Assume that for a given index i, $1 \le i < n$, for the system

$$\dot{z}_1 = z_2 + \sum_{j=1}^{p} \theta_j(t)\phi_{1,j}(z_1),$$
$$\vdots \qquad (44)$$
$$\dot{z}_i = v_i + \sum_{j=1}^{p} \theta_j(t)\phi_{i,j}(z_1, \ldots, z_i),$$

there exists a filtered transformation

$$\bar{z}_1 = z_1,$$
$$\bar{z}_j = z_j - z_j^*(z_1, \ldots, z_{j-1}, \hat{\mu}_1, \ldots, \hat{\mu}_{j-1}),$$
$$z_j^*(0, \ldots, 0, \hat{\mu}_1, \ldots, \hat{\mu}_{j-1}) = 0, \quad 2 \le j \le i,$$
$$\dot{\hat{\mu}}_j = \eta_j(z_1, \ldots, z_j, \hat{\mu}_1, \ldots, \hat{\mu}_{j-1}),$$
$$\eta_j(0, \ldots, 0, \hat{\mu}_1, \ldots, \hat{\mu}_{j-1}) = 0, \quad 2 \le j \le i-1, \qquad (45)$$

and a self-tuning control

$$v_i = z_{i+1}^*(z_1, \ldots, z_i, \hat{\mu}_1, \ldots, \hat{\mu}_i),$$
$$z_{i+1}^*(0, \ldots, 0, \hat{\mu}_1, \ldots, \hat{\mu}_i) = 0,$$
$$\dot{\hat{\mu}}_i = \eta_i(z_1, \ldots, z_i, \hat{\mu}_1, \ldots, \hat{\mu}_{i-1}),$$
$$\eta_i(0, \ldots, 0, \hat{\mu}_1, \ldots, \hat{\mu}_{i-1}) = 0, \qquad (46)$$

such that for the closed loop system (44), (46) the time derivative of the function

$$V_i = \tfrac{1}{2} \sum_{j=1}^{i} (\bar{z}_j^2 + \tilde{\mu}_j^2), \qquad (47)$$

with $\tilde{\mu}_j = \mu_j - \hat{\mu}_j$, μ_j suitable constants, satisfies the inequality

$$\dot{V}_i \le -\sum_{j=1}^{i} (k_j - i + 1)\bar{z}_j^2, \qquad (48)$$

with $k_j - i > 0$, $1 \le j \le i$. Then, for the system

$$\dot{z}_1 = z_2 + \sum_{j=1}^{p} \theta_j(t)\phi_{1,j}(z_1),$$
$$\vdots \qquad (49)$$
$$\dot{z}_{i+1} = v_{i+1} + \sum_{j=1}^{p} \theta_j(t)\phi_{i+1,j}(z_1, \ldots, z_{i+1}),$$

there exists a self-tuning control

$$v_{i+1} = z_{i+2}^*(z_1, \ldots, z_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_{i+1}),$$
$$z_{i+2}^*(0, \ldots, 0, \hat{\mu}_1, \ldots, \mu_{i+1}) = 0,$$
$$\dot{\hat{\mu}}_{i+1} = \eta_{i+1}(z_1, \ldots, z_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i),$$
$$\eta_{i+1}(0, \ldots, 0, \hat{\mu}_1, \ldots, \hat{\mu}_i) = 0, \qquad (50)$$

such that for the closed loop system (49), (50) in the new coordinates given by the filtered transformation

$$\bar{z}_1 = z_1,$$
$$\bar{z}_j = z_j - z_j^*(z_1, \ldots, z_{j-1}, \hat{\mu}_1, \ldots, \hat{\mu}_{j-1}),$$
$$2 \le j \le i+1, \qquad (51)$$

the time derivative of the function

$$V_{i+1} = \tfrac{1}{2} \sum_{j=1}^{i+1} (\bar{z}_j^2 + \tilde{\mu}_j^2)$$
$$= V_i + \tfrac{1}{2}(\bar{z}_{i+1}^2 + \tilde{\mu}_{i+1}^2), \qquad (52)$$

with $\tilde{\mu}_{i+1} = \mu_{i+1} - \hat{\mu}_{i+1}$, μ_{i+1} a suitable constant, satisfies the inequality

$$\dot{V}_{i+1} \le -\sum_{j=1}^{i+1} (k_j - i)\bar{z}_j^2, \qquad (53)$$

with $k_{i+1} - i > 0$.

Proof of the claim. Consider the function

$$\bar{V}_{i+1} = V_i + \tfrac{1}{2}\bar{z}_{i+1}^2. \qquad (54)$$

In view of (44), (47), (48) and (51), we have

$$\dot{V}_{i+1} \le -\sum_{j=1}^{i}(k_j - i + 1)\bar{z}_j^2 + \bar{z}_i\bar{z}_{i+1} + \bar{z}_{i+1}\dot{\bar{z}}_{i+1}. \quad (55)$$

From (45) and (49), we obtain

$$\dot{\bar{z}}_{i+1} = v_{i+1} + \sum_{j=1}^{p}\theta_j(t)\phi_{i+1,j}(z_1, \ldots, z_{i+1}) - \sum_{j=1}^{i}\left(\frac{\partial z_{i+1}^*}{\partial z_j}\dot{z}_j + \frac{\partial z_{i+1}^*}{\partial \hat{\mu}_j}\dot{\hat{\mu}}_j\right). \quad (56)$$

Since $\phi_{j,k}(0, \ldots, 0) = 0$, $1 \le j \le i+1$, $1 \le k \le p$, we can write (see Nijmeijer and van der Scheft (1990))

$$\phi_{j,k}(z_1, \ldots, z_j) = \sum_{l=1}^{j}\bar{\phi}_{l,k}(z_1, \ldots, z_j)z_l, \quad (57)$$

which by (51) implies that we also can write

$$\phi_{j,k}(z_1, \ldots, z_j) = \sum_{l=1}^{j}\bar{\phi}_{l,k}(\bar{z}_1, \ldots, \bar{z}_j)\bar{z}_l. \quad (58)$$

Therefore, we have

$$\begin{aligned}\sum_{j=1}^{p}\theta_j(t)\phi_{i+1,j}(z_1, \ldots, z_j) &- \sum_{j=1}^{i}\frac{\partial z_{i+1}^*}{\partial z_j} \times \sum_{j=1}^{p}\theta_k(t)\phi_{j,k}(z_1, \ldots, z_j)\\ &= \sum_{j=1}^{p}\theta_j(t)\sum_{l=1}^{i+1}\bar{\phi}_{l,j}(\bar{z}_1, \ldots, \bar{z}_{i+1})\bar{z}_l - \sum_{j=1}^{i}\frac{\partial z_{i+1}^*}{\partial z_j}\sum_{k=1}^{p}\theta_k(t)\sum_{l=1}^{j}\bar{\phi}_{l,k}(\bar{z}_1, \ldots, \bar{z}_j)\bar{z}_l\\ &\triangleq \sum_{j=1}^{p}\theta_j(t)\sum_{l=1}^{i+1}\psi_{j,l}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i)\bar{z}_l.\end{aligned}$$

Since $\theta \in \Omega$, a compact set, we can find a smooth function $\alpha_{i+1}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i)$ and an unknown positive constant μ_{i+1} such that

$$\begin{aligned}|\psi_{j,l}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i)| &\le \alpha_{i+1}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i), \\ &\quad 1 \le j \le p, \quad 1 \le l \le i+1, \\ |\theta_j(t)| &\le \frac{1}{p}\sqrt{\mu_{i+1}}, \quad 1 \le j \le p,\end{aligned} \quad (59)$$

so that

$$\left|\sum_{j=1}^{p}\theta_j(t)\sum_{l=1}^{i+1}\psi_{j,l}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i)\right| \le \sqrt{\mu_{i+1}}\,\alpha_{i+1}(\xi_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_i). \quad (60)$$

Define v_{i+1} as

$$\begin{aligned}v_{i+1} &= -\bar{z}_i - (k_{i+1} - i + 1)\bar{z}_{i+1} - \hat{\mu}_{i+1}\bar{z}_{i+1}\alpha_{i+1}^2 + \sum_{j=1}^{i}\left(\frac{\partial z_{i+1}^*}{\partial \hat{\mu}_j}\dot{\hat{\mu}}_j + \frac{\partial z_{i+1}^*}{\partial z_j}z_{j+1}\right)\\ &\triangleq z_{i+1}^*(z_1, \ldots, z_{i+1}, \hat{\mu}_1, \ldots, \hat{\mu}_{i+1}),\end{aligned} \quad (61)$$

with

$$\dot{\hat{\mu}}_{i+1} = \bar{z}_{i+1}^2\alpha_{i+1}^2, \quad (62)$$

and $k_{i+1} - 1 > 0$. Consider the function ($\tilde{\mu}_{i+1} = \mu_{i+1} - \hat{\mu}_{i+1}$)

$$V_{i+1} = \bar{V}_{i+1} + \tfrac{1}{2}\tilde{\mu}_{i+1}^2. \quad (63)$$

Taking (55), (56), (61) and (62) into account, we obtain

$$\begin{aligned}\dot{V}_{i+1} &\le -\sum_{j=1}^{i+1}(k_j - i + 1)\bar{z}_j^2 - \mu_{i+1}\bar{z}_{i+1}^2\alpha_{i+1}^2 + |\bar{z}_{i+1}|\,\|\xi_{i+1}\|\,\alpha_{i+1}\sqrt{\mu_{i+1}}\\ &\le -\sum_{j=1}^{i+1}(k_j - i)\bar{z}_j^2 - \begin{bmatrix}\|\xi_{i+1}\|\\ |\bar{z}_{i+1}|\end{bmatrix}^T \times \begin{bmatrix}1 & -\dfrac{\sqrt{\mu_{i+1}}\,\alpha_{i+1}}{2}\\ -\dfrac{\sqrt{\mu_{i+1}}\,\alpha_{i+1}}{2} & \mu_{i+1}\alpha_{i+1}^2\end{bmatrix} \times \begin{bmatrix}\|\xi_{i+1}\|\\ |\bar{z}_{i+1}|\end{bmatrix}\\ &\le -\sum_{j=1}^{i+1}(k_j - 1)\bar{z}_j^2.\end{aligned} \quad (64)$$

To prove that the claim holds for $i = 1$, we consider the function

$$v_1 = -(k_1 - 1)z_1 - \hat{\mu}_1 z_1\alpha_1(z_1) \triangleq z_2^*(z_1, \hat{\mu}_1), \quad (65)$$

in which $k_1 - 1 > 0$ and $\alpha_1(z_1)$ is a smooth function such that

$$\left|\sum_{j=1}^{p}\theta_j(t)\psi_{1,j}(z_1)\right| \le \mu_1\alpha_1(z_1), \quad (66)$$

with μ_1 a suitable unknown positive real and

$$\psi_{1,j}(z_1) = \frac{\phi_{1,j}(z_1)}{z_1}, \quad 1 \le j \le p,$$

($\psi_{1,j}$ is continuous since $\phi_{1,j}(0) = 0$). The dynamics of $\hat{\mu}_1$ are chosen as

$$\dot{\hat{\mu}}_1 = z_1^2\alpha_1(z_1). \quad (67)$$

It is easy to verify that the function

$$V_1 = \tfrac{1}{2}z_1^2 + \tfrac{1}{2}\tilde{\mu}_1^2, \quad (68)$$

with $\tilde{\mu}_1 = \mu_1 - \hat{\mu}_1$, has time derivative such that

$$\dot{V}_1 \le -k_1 z_1^2. \quad (69)$$

Applying $(n-1)$-times the claim we can construct a function $v_n = z_{n+1}^*(z_1, \ldots, z_n, \hat{\mu}_1, \ldots, \hat{\mu}_n)$ which determines the control

$$v = v_n, \quad (70)$$

and a change of coordinates

$$z = (z_1, \ldots, z_n) \to \bar{z} = (\bar{z}_1, \ldots, \bar{z}_n), \quad (71)$$

such that the function

$$V_n = \frac{1}{2} \sum^{n} (\tilde{z}_i^2 + \tilde{\mu}_i^2) \triangleq V, \tag{72}$$

has time derivative satisfying the inequality

$$\dot{V} \leq -c\, \|\tilde{z}\|^2, \tag{73}$$

with $c > 0$. Consequently, $\|\tilde{z}(t)\|$ and the estimates $\hat{\mu}_i(t)$ are bounded, which imply that $\|z(t)\|$ is bounded. The control variable u which depends on z and on the estimates $\hat{\mu}_i$ is also bounded and, in turn, $\|\dot{z}(t)\|$ is bounded (see (43)). According to (56) and (61), $\|\dot{\tilde{z}}(t)\|$ is bounded and therefore the function $\tilde{z}^T\tilde{z}$ is uniformly continuous. This fact along with the following inequality

$$\lim_{t\to\infty} c \int_0^t \|\tilde{z}(\tau)\|^2\, d\tau \leq \lim_{t\to\infty} - \int_0^t \dot{V}(\tau)\, d\tau$$
$$= V(0) - \lim_{t\to\infty} V(t) < \infty, \tag{74}$$

imply by Barbalat's Lemma (Popov, 1973) that

$$\lim_{t\to\infty} \|\tilde{z}(t)\| = 0, \tag{75}$$

and, by (5) and (71) that

$$\lim \|x(t)\| = 0. \tag{76}$$

Remark 4.1. Note that the dynamic compensator in Theorem 4.1 is of order at most n and therefore independent of the dimension of the vector θ.

Example 4.1. Consider the system

$$\begin{aligned} \dot{x}_1 &= x_2 + \theta_1(t)x_1^2 + \theta_2(t)x_1^3, \\ \dot{x}_2 &= u, \end{aligned} \tag{77}$$

where $\theta_1(t)$ and $\theta_2(t)$ are bounded disturbances. System (77) satisfies the conditions of Theorem 4.1. Following the proof of Theorem 4.1 we define (see (65))

$$v_1 = -(k_1 - 1)x_1 - \hat{\mu}_1 x_1 \alpha_1(x_1),$$

where $k_1 > 1$ and, according to (66), $\alpha_1(x_1)$ can be chosen as

$$\alpha_1(x_1) = 1 + x_1^2.$$

The dynamics of $\hat{\mu}_1$ are chosen as in (67)

$$\dot{\hat{\mu}}_1 = x_1^2 \alpha_1(x_1).$$

Define, as in (51),

$$\tilde{x}_2 = x_2 - v_1.$$

According to (61) the final control u is

$$u = v_2 = -x_1 - k_2\tilde{x}_2 - \hat{\mu}_2\tilde{x}_2\alpha_2^2(x_1, \hat{\mu}_1) - \dot{\hat{\mu}}_1 x_1 \alpha_1(x_1) - \frac{\partial v_1}{\partial x_1} x_2,$$

with $k_2 > 0$ and

$$\dot{\hat{\mu}}_2 = \tilde{x}_2^2 \alpha_2^2(x_1, \hat{\mu}_1).$$

The function α_2 has to satisfy, for some unknown value of μ_2, the inequality (60) which for system (77) becomes

$$\frac{\partial v_1}{\partial x_1}(\theta_1(t)x_1 + \theta_2(t)x_1^2) \leq \sqrt{\mu_2}\, \alpha_2(x_1, \hat{\mu}_1).$$

A possible choice for α_2 is

$$\alpha_2(x_1, \hat{\mu}_1) = 1 + (1 + \hat{\mu}_1^2)x_1^4.$$

The proof of Theorem 4.1 allows us to generalize Corollary 3.1 and to determine a *fixed* stabilizing controller for a family of linear systems in controller canonical form perturbed by global Lipschitz nonlinearities with unknown Lipschitz constants.

Theorem 4.2. Consider the system

$$\dot{z} = Az + bu + \begin{bmatrix} \phi_1(z, \theta(t)) \\ \vdots \\ \phi_n(z, \theta(t)) \end{bmatrix}, \quad z \in R^n, \tag{78}$$

with (A, b) in controller canonical form (8). If

$$|\phi_i(z, \theta(t))| \leq \mu_i \, \|[z_1, \ldots, z_i]^T\|, \quad 1 \leq i \leq n, \tag{79}$$

with unknown Lipschitz constants μ_i, then there exists a globally stabilizing dynamic state feedback compensator of order n.

Proof. Even though the strict triangularity and the linear parametrization assumptions are not satisfied they are satisfied by the bounding functions in (79) so that we can follow the arguments used in Theorems 4.1. The assumption that Ω is known, which is required in Theorem 3.1, is not needed.

Example 4.2. Consider the system

$$\begin{aligned} \dot{x}_1 &= x_2 + \theta(t)x_1 \sin x_3, \\ \dot{x}_2 &= x_3, \\ \dot{x}_3 &= u, \end{aligned}$$

with $\theta(t)$ belonging to the unknown interval $[\theta_m, \theta_M]$. Since

$$|\theta(t)x_1 \sin x_3| \leq \max(|\theta_m|, |\theta_M|)|x_1|,$$

Theorem 4.2 applies. Note that the strict triangularity assumption is not satisfied.

5. CONCLUSIONS

We have addressed the robust state feedback stabilization problem for single-input, time-varying uncertain nonlinear systems. Assuming

that the nominal time-invariant system is globally feedback linearizable, that the time-varying nonlinear uncertainties satisfy structural coordinate-free triangularity conditions and that the time-varying unknown parameter vector belongs to a known compact set, we construct in Theorem 3.1 a fixed static state feedback control which globally stabilizes each system in the family.

In Theorem 4.1 we show that if the unknown parameter vector enters linearly, the knowledge of the compact set in which it varies is not required. Theorem 4.2 shows that the techniques developed in this paper lead to the construction of a fixed self-tuning control which globally stabilizes any linear system in controller canonical form perturbed by time-varying unknown nonlinearities as long as they are bounded by globally Lipschitz functions satisfying the strict triangularity condition, without requiring the knowledge of the Lipschitz constants.

REFERENCES

Akhrif, O and G L Blankeship (1988) Robust stabilization of feedback linearizable systems *Proc IEEE Conf on Decision and Control*, pp 1714–1719

Bacciotti, A (1992) *Local Stabilizability of Nonlinear Control Systems* World Scientific, Singapore

Barmish, B R, M Corless and G Leitmann (1983) A new class of stabilizing controllers for uncertain dynamical systems *SIAM J Control Optimiz*, **21**, 246–255

Boothby, W M (1984) Some comments on global linearization of nonlinear systems *Systems and Control Letters*, **4**, 143–147

Boothby, W M (1986) Global feedback linearizability of locally linearizable systems In M Fliess and M Hazewinkel (Eds), *Algebraic and Geometric Methods in Nonlinear Control Theory*, pp 243–256 D Reidel, Dordrecht

Dayawansa, W P, W M Boothby and D Elliott (1985). Global state and feedback equivalence of nonlinear systems *Systems and Control Letters*, **6**, 229–234

Hahn, W (1967) *Stability of Motion* Springer-Verlag, Berlin

Isidori, A (1989) *Nonlinear Control Systems* Springer-Verlag, Berlin

Jakubczyk, B and W Respondek (1980) On linearization of control systems *Bull Acad Polonaise Sci Ser Sci Math*, **28**, 517–522

Kanellakopoulos, I, P V Kokotovic and R Marino (1991a) An extended direct scheme for robust adaptive nonlinear control *Automatica*, **27**, 247–255

Kanellakopoulos, I, P V Kokotovic and A S Morse (1991b) Systematic design of adaptive controllers for feedback linearizable systems *IEEE Trans Aut Control*, **AC-36**, 1241–1253

Nam, K and A Arapostathis (1988) A model reference adaptive control scheme for pure-feedback nonlinear systems *IEEE Trans Aut Control*, **AC-33**, 803–811

Nijmeijer, H and A van der Schaft (1990) *Nonlinear Dynamical Control Systems* Springer-Verlag, Berlin

Popov, V M (1973) *Hyperstability of Control Systems* Springer-Verlag, Berlin

Respondek, W (1986) Global aspects of linearization, equivalence to polynomial forms and decomposition of nonlinear control systems In M Fliess and M Hazewinkel (Eds) *Algebraic and Geometric Methods in Nonlinear Control Theory* pp 243–256 D Reidel, Dordrecht

Sastry, S S and A Isidori (1989) Adaptive control of linearizable systems *IEEE Trans Aut Control*, **AC-34**, 1123–1131

Sontag E D (1990) Feedback stabilization of nonlinear systems In M A Kaashoek, J H van Schuppen and A C M Ran (Eds), *Robust Control of Linear Systems and Nonlinear Control*, pp 61–81

Su R (1982) On the linear equivalents of nonlinear systems *Systems and Control Letters* **2**, 48–52

Taylor, D, P V Kokotovic, R Marino and I Kanellakopoulos (1989) Adaptive regulation of nonlinear systems with unmodelled dynamics *IEEE Trans Aut Control*, **34**, 405–412

Thorp, J S and B R Barmish (1981) On guaranteed stability of uncertain linear systems via linear control *J of Opt Theory and Applications*, **35**, 559–579

Automatica Vol 29 No pp 191 200 1993
Printed in Great Britain

0005 1098/93 $6 00 + 0 00

Fading-memory Feedback Systems and Robust Stability*†

JEFF S SHAMMA‡§ and RONGZE ZHAO||

A small-gain condition is necessary for robust stability of fading-memory closed-loop feedback systems with unstructured model uncertainty Linear plants stabilized by fading-memory nonlinear compensators lead to closed-loop fading-memory systems

Key Words—Nonlinear systems, time varying systems robust control robustness system theory

Abstract—This paper considers fading memory for nonlinear time-varying systems and associated problems of robust stability

We define two notions of fading memory for stable dynamical systems uniform and pointwise We then provide conditions under which stable linear or nonlinear systems exhibit uniform or pointwise fading memory In particular we show that (1) all stable discrete time linear time varying (LTV) systems have uniform fading memory (2) all stable continuous time LTV systems have pointwise fading memory and (3) stable finite-dimensional continuous time LTV systems have uniform fading memory

We then show that a version of the small gain theorem which employs the asymptotic gain of a fading memory system is necessary for the stable invertibility of certain feedback operators These results are presented for both continuous time and discrete time systems using general ℓ^p or $\mathcal{L}^p$ notions of input/output stability and generalize existing results for ℓ^2 stability We further investigate fading memory in a closed loop context For linear plants we parametrize all nonlinear controllers which lead to closed loop pointwise fading memory

1 INTRODUCTION

THE PROBLEM OF robust stability analysis (cf Dorato (1987) and references therein) is to determine under what conditions a given controller stabilizes a prescribed *family* of possible plants Typically, this plant family arises from various approximations, simplifications, and limitations in the plant modeling process One framework for plant family representations is that of unstructured uncertainty More precisely, the plant family is represented as a nominal plant combined with a norm-bounded perturbation The theory for robust stability analysis for linear time-invariant systems subject to unstructured uncertainty is well developed (e g Chen and Desoer (1982), Dahleh and Ohta (1988), Doyle (1982), Doyle and Stein (1981), Doyle *et al* (1982), Georgiou and Smith (1990), Khammash and Pearson (1991))

Linear systems typically arise as linearizations of nonlinear systems Furthermore, adaptive control laws for linear systems are typically nonlinear Thus, robust stability analysis tools for nonlinear systems are desirable For nonlinear systems, a standard tool for stability analysis is the small gain theorem (Desoer and Vidyasagar (1975), Sandberg (1965), Zames (1966)) A limitation of the small gain theorem is that it can often be a conservative sufficient condition for stability Recent work by the author (Shamma (1991)) has shown that the small gain theorem is in fact necessary when considering nonlinear plant families characterized by a norm-bounded perturbation This work was developed for ℓ^2 (i e finite-energy) stability of discrete-time systems The nonlinear systems considered in Shamma (1991) are those with *fading memory* (e g Boyd and Chua (1985)) Intuitively, a fading memory property means that the current output depends on the recent inputs and not the remote past Thus, fading-memory is a reasonable assumption for many physical systems

The objective of this paper is to further develop the analysis of fading memory systems

* Received 23 August 1991 revised 3 February 1992 received in final form 22 May 1992 This paper was not presented at any IFAC meeting This paper was recommended for publication in revised form by Guest Editor B A Francis

† Research supported in part by an AFOSR Summer Faculty Fellowship at the USAF Armament Directorate Eglin Air Force Base, FL 32542, AFOSR Research Initiation grant #92 70 subcontracted to the Research and Development Laboratories, NSF Research Initiation grant #ECS 9296074

‡ Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin Austin TX 78712, U S A

§ Formerly with the Department of Electrical Engineering University of Minnesota Minneapolis MN 55455, U S A

|| Control Sciences and Dynamical Systems University of Minnesota, Minneapolis, MN 55455 U S A

in the context of robust stability. The results here are presented in an essentially norm-invariant manner. We consider stability over arbitrary $\mathscr{L}_p$ or ℓ_p spaces with $p \in [1, \infty)$. We also consider a form of bounded-input/bounded-output $\mathscr{L}^\infty$ or ℓ^∞ stability with asymptotic decay. Given the norm-invariant nature of the presentation, it is the fading memory property which is isolated and exploited to lead to the desired results.

In this paper, we define two notions of fading memory for stable dynamical systems: uniform and pointwise. In uniform fading memory, the effects of a finite-duration of input eventually vanish depending on the length of the duration only. In pointwise fading memory, the effects of a finite duration of input also eventually vanish but now depending on the length of the duration and the input itself.

The distinction between uniform and pointwise fading memory turns out to be important. In particular, we will see that all stable discrete-time linear systems exhibit uniform fading memory, all stable continuous-time linear systems exhibit pointwise fading memory, and stable finite-dimensional continuous-time linear systems exhibit uniform fading memory. The pointwise fading memory property also allows us to weaken the conditions under which the small gain theorem is necessary. This leads to a larger class of nonlinear systems for which these results are applicable.

The fading memory property is also considered in a closed-loop context. More precisely, we consider under what conditions a stabilizing compensator leads to a closed-loop system with pointwise fading memory. A key tool in this analysis is the factorization representation for nonlinear operators (cf. Verma (1988) and references therein). This allows us to parametrize all nonlinear compensators for linear plants which lead to closed-loop stability with pointwise fading memory. This parametrization takes the form of the familiar linear fractional parametrization (e.g. Francis (1987), Youla *et al.* (1976)) with the free parameter having fading memory.

The remainder of this paper is organized as follows. In Section 2, we establish notation and state some preliminary results. In Section 3, we consider the fading memory property. In Section 3.1, we define uniform and pointwise fading memory. In Section 3.2, we concentrate on the differences between pointwise and uniform fading memory and provide conditions for an operator to have pointwise fading memory. In Section 4, we generalize previous results (Shamma (1991)) which show that the small gain theorem is necessary for robust stability of fading memory nonlinear systems. These results are applied to a robust stabilization problem in Section 5. Section 5 also presents conditions under which a closed-loop system exhibits fading memory. Finally, Section 6 contains some concluding remarks.

2 MATHEMATICAL PRELIMINARIES

In this paper, we will consider both discrete-time and continuous-time systems with several notions of finite-gain stability (cf. Desoer and Vidyasagar (1975); Willems (1971)). Towards this end, the symbol $\mathscr{V}$ is used to denote any *one* of the following normed signal spaces: ℓ^p or $\mathscr{L}^p$ with $p \in [1, \infty)$, $\mathscr{L}_0^\infty$, and c_0, where

$$\mathscr{L}_0^\infty \stackrel{\text{def}}{=} \left\{ f \in \mathscr{L}^\infty : \lim_{t\to\infty} \sup_{\tau > t} |f(\tau)| = 0 \right\},$$

$$c_0 \stackrel{\text{def}}{=} \left\{ f \in \ell^\infty : \lim |f(n)| = 0 \right\}.$$

The spaces $\mathscr{L}^p$ and ℓ^p are equipped with the usual norms, all denoted $\|\cdot\|$. The spaces $\mathscr{L}_0^\infty$ and c_0 are equipped with the usual "supremum" norms, also denoted $\|\cdot\|$. The symbol $\mathscr{T}$ is used to denote either $\mathscr{R}^+$ or $\mathscr{Z}^+$. Occasionally, it will be necessary to specify or restrict the particular definition of $\mathscr{V}$ and $\mathscr{T}$.

Let $f : \mathscr{T} \to \mathscr{R}^n$. The support of f is denoted supp(f). The restriction of f to the interval $[a, b]$ is denoted $f|_{[a,b]}$. For $T \in \mathscr{T}$, S_T denotes the T-shift (time-delay) operator:

$$S_T f(t) \stackrel{\text{def}}{=} \begin{cases} 0, & t < T; \\ f(t - T), & t \geq T, \end{cases}$$

and P_T denotes the truncation operator:

$$P_T f(t) \stackrel{\text{def}}{=} \begin{cases} f(t), & t \leq T; \\ 0, & t > T. \end{cases}$$

The extended space, $\mathscr{V}_e$, is defined as

$$\mathscr{V}_e \stackrel{\text{def}}{=} \{ f : \mathscr{T} \to \mathscr{R}^n : P_T f \in \mathscr{V}, \forall T \in \mathscr{T} \}.$$

The set of all $f \in \mathscr{V}_e$ with $f \notin \mathscr{V}$ is denoted by $\mathscr{V}_e \backslash \mathscr{V}$.

Let $H : \mathscr{V}_e \to \mathscr{V}_e$. Then H is called *causal* if

$$P_T H f = P_T H P_T f, \quad \forall T \in \mathscr{T},$$

time-invariant if

$$H S_T = S_T H, \quad \forall T \in \mathscr{T},$$

stable if $f \in \mathscr{V}$ implies $Hf \in \mathscr{V}$ with

$$\|H\| \stackrel{\text{def}}{=} \sup_{\substack{f \in \mathscr{V} \\ f \neq 0}} \frac{\|Hf\|}{\|f\|} < \infty,$$

and *incrementally stable* if it is stable with

$$\|H\|_\delta \stackrel{\text{def}}{=} \sup_{\substack{f_1, f_2 \in \mathcal{V} \\ f_1 \neq f_2}} \frac{\|Hf_1 - Hf_2\|}{\|f_1 - f_2\|} < \infty.$$

The definitions of stability over $\mathcal{L}_0^\infty$ and c_0 are somewhat non-standard. For causal operators, stability over $\mathcal{L}_0^\infty$ implies stability over $\mathcal{L}^\infty$ with the same induced norm. The main difference is that stability over $\mathcal{L}_0^\infty$ implies a notion of asymptotic stability in a bounded-input/bounded-output setting. Similar arguments hold for stability over c_0.

Henceforth, all operators are assumed to be causal. The set of all *stable* NLTV operators $H: \mathcal{V}_e \to \mathcal{V}_e$ is denoted $\mathcal{S}_{Nl}$. The subset of operators in $\mathcal{S}_{Nl}$ which are *linear* (and possibly time varying) is denoted $\mathcal{S}_l$.

The following definition is adapted from Willems (1971).

Definition 2.1. The operator $I - G: \mathcal{V}_e \to \mathcal{V}_e$ is said to be *causally invertible* if
(1) $I - G$ is one-to-one and onto
(2) $(I - G)^{-1}$ is causal.

A sufficient condition for causal invertibility of $I - G$ (cf. Willems (1971)) is that G can be factored as $G = S_\epsilon \tilde{G}$ or $G = \tilde{G} S_\epsilon$, where $\epsilon > 0$.

Finally a preliminary lemma is presented.

Lemma 2.1. Let $v, w \in \mathcal{V}$ with $\operatorname{supp}(v) = [T_1, T_2]$ and $\operatorname{supp}(w) = [T_3, T_4]$ with $T_3 > T_2$. There exists a $G \in \mathcal{S}_l$ such that:
(1) $w = Gv$.
(2) $\operatorname{supp}(Gf) \subset [T_3, T_4]$, $\forall f \in \mathcal{V}$.
(3) $\operatorname{supp}(f) \subset \mathcal{T} - [T_1, T_2] \Rightarrow Gf = 0$.
(4) $\|G\| = \|w\|$.

Proof. For simplicity, the proof is given for the single-input/single-output case only.

It will be necessary to distinguish between the various definitions of $\mathcal{V}$ and $\mathcal{T}$. First, let $\mathcal{V}$ denote $\mathcal{L}^p$ for some $p \in [1, \infty)$, and let $\mathcal{T}$ denote $\mathcal{R}^+$. Define $g(\cdot, \cdot): \mathcal{V} \times \mathcal{V} \to \mathcal{R}$ by

$$g(t, \tau) \stackrel{\text{def}}{=} \frac{1}{\|v\|} \; w(t)\,|v(\tau)|^{p-1} \operatorname{sign}(v(\tau)).$$

Now define $G \in \mathcal{S}_L$ by

$$(Gf)(t) \stackrel{\text{def}}{=} \int_0^t g(t, \tau) f(\tau)\, d\tau.$$

Condition 1 follows immediately from this definition. Using that $g(t, \tau) = 0$ for $\tau \notin [T_1, T_2]$ or $t \notin [T_3, T_4]$ leads to Conditions 2–3. Condition 4 follows from a straightforward application of the Hölder inequality (e.g. Desoer and Vidyasagar (1975)).

Now let $\mathcal{V}$ denote $\mathcal{L}_0^\infty$. Since $\operatorname{supp}(v) = [T_1, T_2]$, we can consider v as an element of $\mathcal{L}^\infty[T_1, T_2]$. From the Hahn–Banach theorem (Rudin (1987)) there exists a $z \in (\mathcal{L}^\infty[T_1, T_2])^*$ such that $\|z\| = 1$ and $z(v) = \|v\|$. Then define G by

$$(Gf)(t) = \frac{1}{\|v\|} w(t) z(f\,|_{[T_1, T_2]}).$$

Note that since $w(t) = 0$ for $t < T_3$, this operator is causal. Conditions 1–4 follow immediately from this definition. In fact, the construction in case $\mathcal{V}$ denotes $\mathcal{L}^p$ follows these lines with the dual element z stated explicitly.

The construction for discrete-time ($\mathcal{V}$ denotes ℓ^p or c_0, and $\mathcal{T}$ denotes z^+) follows from similar arguments.

3 FADING-MEMORY

3.1. *Definitions and conditions for fading memory*

A notion of fading-memory is the central focus of this paper. Intuitively, fading-memory implies that the effects of a finite-duration of input eventually vanish. A definition of fading-memory tailored to stable time-invariant systems over ℓ^∞ or $\mathcal{L}^\infty$ signals was given in Boyd and Chua (1985). An alternate definition for stable time-varying systems over ℓ^2 was given in Shamma (1991). The following definitions generalize those in Shamma (1991) to stable time-varying systems over $\mathcal{V}$.

Definition 3.1. An operator $H \in \mathcal{S}_{Nl}$ is said to have *pointwise finite-memory* if there exists a function $FM(\cdot, \cdot, H): \mathcal{V} \times \mathcal{T} \to \mathcal{T}$ such that for all $f \in \mathcal{V}$ and $t \in \mathcal{T}$
(1) $FM(f, t; H) \geq t$,
(2) $FM(f, t; H) = FM(P_t f, t; H)$,
(3) $(I - P_{FM(f,t,H)})Hf = (I - P_{FM(f,t,H)})H(I - P_t)f$.
Note that $FM(f, t; H)$ is causally dependent on f.

This definition of pointwise finite-memory is somewhat weaker than that in Shamma (1991). The definition in Shamma (1991) requires that inputs over a given finite-duration are forgotten *uniformly* as follows.

Definition 3.2. An operator $H \in \mathcal{S}_{Nl}$ is said to have *uniform finite-memory* if there exists a function $FM(\cdot; H): \mathcal{T} \to \mathcal{T}$ such that for all $f \in \mathcal{V}$ and $t \in \mathcal{T}$
(1) $FM(t; H) \geq t$,
(2) $(I - P_{FM(t,H)})Hf = (I - P_{FM(t,H)})H(I - P_t)f$.

The notation $FM(\cdot)$ is used both for uniform fading-memory and pointwise fading-memory. From the different number of arguments, this relaxed notation should not lead to any confusion.

Fading-memory operators are now defined as follows.

Definition 3.3. An operator $H \in \mathscr{S}_{NL}$ is said to have *uniform (resp. pointwise) fading-memory* if it can be approximated arbitrarily closely in norm by uniform (resp. pointwise) finite-memory operators.

Some useful consequences of these definitions are now presented. The proofs are slight modifications of similar propositions in Shamma (1991) and are omitted.

The following proposition states that pointwise finite-memory nonlinear operators are right-distributive over signals with sufficient time-separation.

Proposition 3.1. Let H have pointwise finite-memory. Let $f_1 \in \mathscr{V}$ have $\operatorname{supp}(f_1) \subset [0, T_1]$. Let $T_2 = FM(f_1, T_1; H)$. Then for all $f_2 \in \mathscr{V}$ with $\operatorname{supp}(f_2) \subset (T_2, \infty)$,

$$H(f_1 + f_2) = Hf_1 + Hf_2.$$

The following theorem characterizes how the effects of a finite-duration input eventually decay in uniform fading-memory operators.

Theorem 3.1. For $H \in \mathscr{S}_{NL}$, the following statements are equivalent:
(a) The operator H has uniform fading-memory.
(b) Given any $\varepsilon > 0$, there exists a function $FM_\varepsilon(\cdot\,; H): \mathscr{T} \to \mathscr{T}$ with $FM_\varepsilon(t; H) \geq t$ such that

$$\|(I - P_{FM_\varepsilon(t,H)})Hf - (I - P_{FM_\varepsilon(t,H)})H(I - P_t)f\| \leq \varepsilon\, \|f\|,$$

for all $f \in \mathscr{V}$ and $t \in \mathscr{T}$.

The following theorem states that all discrete-time linear operators have uniform fading-memory.

Theorem 3.2. Let $\mathscr{V}$ denote ℓ^p with $p \in [1, \infty)$ or c_0. Then every $H \in S_L$ has uniform fading-memory.

Some examples illustrating fading-memory are the following.

Example 3.1. A linear operator over ℓ^∞ without pointwise fading-memory.

Define $H: \ell^\infty \to \ell^\infty$ as $(Hf)(n) = f(0)$.

See Boyd and Chua (1985) for examples of linear time-invariant operators on ℓ^∞ and $\mathscr{L}^\infty$ which do not exhibit fading-memory.

Example 3.2. A nonlinear time-invariant operator without pointwise fading memory.

Define $H: \ell^2 \to \ell^2$ as

$$(Hf)(n) = \operatorname{sat}(\|P_n f\|)f(n),$$

where $\operatorname{sat}(\cdot)$ denotes the standard unity-saturation function.

3.2. *Pointwise fading-memory*

In this section, we will highlight the difference between pointwise and uniform fading memory.

Example 3.3. A pointwise but non-uniform fading-memory operator.

Let $\mathscr{M}$ denote the set of $f \in \mathscr{L}^2[0, 1)$ such that $t \mapsto f\left(\frac{t-1}{t}\right) \in \mathscr{L}^2[1, \infty)$. Define $H: \mathscr{L}^2 \to \mathscr{L}^2$ by

$$(Hf)(t) = \begin{cases} 0, & 0 \leq t < 1; \\ 0, & t \geq 1, f|_{[0,1)} \notin \mathscr{M}; \\ f\left(\frac{t-1}{t}\right), & t \geq 1, f|_{[0,1)} \in \mathscr{M}. \end{cases}$$

In words, the operator H takes the input over the time-interval $[0, 1)$ and stretches it over the time-interval $[1, \infty)$.

From Theorem 3.1, it is clear that H does not have uniform fading-memory. However, H does have *pointwise* fading-memory. To see this, we will construct a pointwise finite-memory approximant, $\tilde{H}$, of H as follows. Given any $\varepsilon > 0$, define $\tilde{H}$ by

$$(\tilde{H}f)(t) = \begin{cases} 0, & 0 \leq t < 1; \\ 0, & t \geq 1, f|_{[0,1)} \notin \mathscr{M}; \\ f\left(\frac{t-1}{t}\right), & t \geq 1, f|_{[0,1)} \in \mathscr{M}, \\ & \quad \|P_{(t-1)/t}f\| \leq (1-\varepsilon)\, \|f|_{[0,1)}\|; \\ 0, & \quad \|P_{(t-1)/t}f\| > (1-\varepsilon)\, \|f|_{[0,1)}\|. \end{cases}$$

From this definition, it is easy to see that $\|H - \tilde{H}\| \leq \varepsilon$. The main idea is that for any signal f, $\tilde{H}$ truncates the portion in the interval $[0, 1)$ which will be stretched to $[1, \infty)$. However, the portion of $[0, 1)$ to be truncated is *signal-dependent*, hence the restriction to pointwise fading-memory only.

The above construction of a pointwise fading-memory approximant provides the main insight for the following theorem.

Theorem 3.3. For $H \in \mathcal{S}_{NL}$, the following statements are equivalent.

(a) The operator H has pointwise fading-memory.

(b) Given any $\varepsilon > 0$ the following condition holds. There exists a function $FM_\varepsilon(\cdot, \cdot; H): \mathcal{V} \times \mathcal{T} \to \mathcal{T}$ such that for all $f \in \mathcal{V}$ and $t \in \mathcal{T}$

(1) $FM_\varepsilon(f, t; H) \geq t$,

(2) $FM_\varepsilon(f, t; H) = FM_\varepsilon(P_t f, t; H)$,

(3) $\|(I - P_{FM_\varepsilon(f,t,H)})Hf - (I - P_{FM_\varepsilon(f,t,H)}) H(I - P_t)f\| \leq \varepsilon \|f\|$.

Proof. $(a \Rightarrow b)$ Given any $\varepsilon > 0$, choose $\bar{H} \in \mathcal{S}_{NL}$ with pointwise finite-memory such that $\|H - \bar{H}\| \leq \varepsilon/2$. Then set the desired $FM_\varepsilon(\cdot, \cdot; H) = FM(\cdot, \cdot; \bar{H})$. To see this, for any $f \in \mathcal{V}$ and $T \in \mathcal{T}$,

$$
\begin{aligned}
&(I - P_{FM_\varepsilon(f,T,H)})Hf \\
&\quad - (I - P_{FM_\varepsilon(f,T,H)})H(I - P_T)f \\
&= (I - P_{FM(f,T,\bar{H})})\bar{H}f \\
&\quad - (I - P_{FM(f,T,\bar{H})})\bar{H}(I - P_T)f \\
&\quad + (I - P_{FM_\varepsilon(f,T,H)})(H - \bar{H})f \\
&\quad - (I - P_{FM_\varepsilon(f,T,H)})(H - \bar{H})(I - P_T)f
\end{aligned}
$$

However,

$$
\begin{aligned}
&(I - P_{FM(f,T,H)})\bar{H}f \\
&\qquad - (I - P_{FM(f,T,H)})\bar{H}(I - P_T)f = 0.
\end{aligned}
$$

Standard norm bounding leads to the desired result.

$(b \Rightarrow a)$ Given $\varepsilon > 0$, we will explicitly construct a pointwise finite-memory approximant, $\bar{H}$, to H. Let $\alpha_n = \varepsilon(\frac{1}{2})^{n+1}$, and let $FM_{\alpha_n}(\cdot, \cdot; H)$ be functions as in condition (b). For any $f \in \mathcal{V}$, define the sequence $\{t_i\} \subset \mathcal{T}$ as

$$
\begin{aligned}
t_0 &= 0 \\
t_1 &= FM_{\alpha_0}(f, t_0; H) + 1 \\
t_2 &= FM_{\alpha_1}(f, t_1; H) + 1 \\
&\vdots
\end{aligned}
$$

Note that for a given ε, this sequence is *dependent* on the particular input, f, but in a *causal* manner. Then set

$$
(\bar{H}f)(t) = \begin{cases} (Hf)(t), & t_0 \leq t < t_1; \\ (H(I - P_{t_0})f)(t), & t_1 \leq t < t_2; \\ (H(I - P_{t_1})f)(t), & t_2 \leq t < t_3; \\ \text{etc.} & \end{cases}
$$

It follows that $\bar{H} \in \mathcal{S}_{NL}$ has pointwise finite-memory. Furthermore,

$$
\begin{aligned}
\|Hf - \bar{H}f\| &\leq \sum_{j=0}^{\infty} \|(Hf - \bar{H}f)|_{[t_j, t_{j+1})}\| \\
&\leq \sum_{j=0}^{\infty} \alpha_j \|f\| \leq \varepsilon \|f\|. \qquad \blacksquare
\end{aligned}
$$

Theorem 3.3 leads to the following corollary that all continuous-time linear systems exhibit pointwise fading-memory.

Corollary 3.1. Let $\mathcal{V}$ denote $\mathcal{L}^p$ with $p \in [1, \infty)$ or $\mathcal{L}_0^\infty$. Then every $H \in S_l$ has pointwise fading-memory.

Finite-dimensional linear systems, however, always exhibit uniform fading-memory.

Theorem 3.4. The input to-state mapping of a continuous-time linear system has uniform fading-memory.

Proof. Let H denote the input-to-state mapping of a continuous-time finite-dimensional linear system stable over $\mathcal{V}$,

$$\dot{g}(t) = A(t)g(t) + f(t).$$

Let $\Phi(t, \tau)$ denote the associated state-transition matrix. Since H is finite-dimensional, Φ admits the decomposition $\Phi(t, \tau) = M(t)M^{-1}(\tau)$ for an appropriate M.

It suffices to construct a function $FM_\varepsilon(\cdot; H)$ as in Theorem 3.1. Since H is linear, the desired condition from Theorem 3.1 becomes

$$\|(I - P_{FM_\varepsilon(T,H)})HP_T\| < \varepsilon.$$

Towards this end, let $f \in \mathcal{V}$ have $\operatorname{supp}(f) \subset [0, T_1]$, and let $g = Hf$. Then for $t > T_1$,

$$g(t) = M(t)\int_0^{T_1} M^{-1}(\tau)f(\tau)\, d\tau.$$

Thus

$$|g(t)| \leq |M(t)|\, \|M^{-1}|_{[0,T_1]}\|_{\mathcal{L}^q}$$

where

$$\frac{1}{p} + \frac{1}{q} = 1,$$

in case $\mathcal{V}$ denotes $\mathcal{L}^p$, $q = 1$ in case $\mathcal{V}$ denotes $\mathcal{L}_0^\infty$, $|M(t)|$ denotes the pertinent induced matrix norm, and

$$\|M^{-1}|_{[0,T_1]}\|_{\mathcal{L}^q} = \left(\int_0^{T_1} |M^{-1}(\tau)|^q\, d\tau\right)^{1/q}.$$

The above relationship can be used to define the desired function FM_ε; namely set $FM_\varepsilon(T_1; H)$ to be the smallest time $T_2 \geq T_1$ such that

$$\|M|_{[T_2,\infty)}\|_{\mathcal{V}}\, \|M^{-1}|_{[0,T_1]}\|_{\mathcal{L}^q} < \varepsilon.$$

4 CONDITIONS FOR ROBUST INVERTIBILITY

The main results of Shamma (1991) deal with uniform fading-memory systems over ℓ^2. In this section, these results are extended to encompass

pointwise fading-memory continuous-time and discrete-time systems over any signal-space in $\mathcal{V}$.

The main results are the following.

Theorem 4.1. Let $\mathcal{V}$ denote any one of the $\mathcal{L}_p$ or ℓ_p spaces with $p \in [1, \infty)$. For $H \in \mathcal{S}_L$, let $Q \in \mathcal{S}_L$ be called "admissible" if the operator $I - QH$ is causally invertible and $\|Q\| < 1$. Then the operator $I - QH$ has a *stable* inverse for all admissible $Q \in \mathcal{S}_L$ if and only if

$$\inf_{t \in \mathcal{T}} \|HS_t\| \leq 1.$$

Definition 4.1. Let $H \in \mathcal{S}_{NL}$ satisfy

$$\inf_{t \in \mathcal{T}} \|HS_t\| = \gamma.$$

Then H is said to have the *uniformity in norm-excitation* (UINE) property if the following condition holds. Given any $\varepsilon \in (0, \gamma)$, there exists a sequence $\{f_k\} \subset \mathcal{V}$ such that

(1) $\|HS_k f_k\| \geq (\gamma - \varepsilon)\|f_k\|, \forall k \in \mathcal{Z}^+$.

(2) $\sup_k \|f_k\| < \infty$.

(3) $\inf_k \|f_k\| > 0$.

The UINE property states that approaching the operator gain, $\|HS_k\|$, does not require injecting signals that become arbitrarily large or small as $k \to \infty$. Note that by the homogeneity property of linear systems, any $H \in \mathcal{S}_L$ has the UINE property. Furthermore, any nonlinear $H = \mathcal{S}_{NL}$ which is *time-invariant* also has the UINE property.

Theorem 4.2. Let $\mathcal{V}$ denote any one of the $\mathcal{L}_p$ or ℓ_p spaces with $p \in [1, \infty)$. Let $H \in \mathcal{S}_{NL}$ have pointwise fading-memory and the UINE property. Let $Q \in \mathcal{S}_{NL}$ be called "admissible" if the operator $(I - QH)$ is causally invertible, Q has pointwise fading-memory, and $\|Q\| < 1$. Then the operator $I - QH$ has a *stable* inverse for all admissible $Q \in \mathcal{S}_{NL}$ and

$$\sup_{Q \text{ admissible}} \|(I - QH)^{-1}\| < \infty,$$

only if

$$\inf_{t \in \mathcal{T}} \|HS_t\| < 1.$$

A key difference between Theorem 4.1 and Theorem 4.2 is that Theorem 4.2 provides necessary conditions for uniform robust invertibility. The reason for this is that in case H is nonlinear (as in Theorem 4.2) a *destabilizing* Q is constructed for H with finite memory but *not* fading memory. Rather, when H has fading memory, it is the uniformity in stability which is violated.

It is important to note that the asymptotic small gain condition of Theorem 4.2, while sufficient for robust invertibility, *is not* sufficient for uniform robust invertibility. It is easy to show (using simply time-varying gains) that the uniform bound in norm can be violated in finite-time. A claim without proof of uniform robust invertibility was made in Shamma (1991). Hence, the statement of Theorem 3.2 in Shamma (1991) should be modified accordingly.

Proof. The following is the necessity proof of Theorem 4.2 for H with pointwise *finite-memory*. This will sufficiently demonstrate the main ideas so that one may modify the proofs in Shamma (1991) appropriately. The main difference between the present results and those of Shamma (1991) is that pointwise finite-memory is assumed rather than the stronger uniform fading-memory.

First, suppose that

$$\inf_{t \in \mathcal{T}} \|HS_t\| \geq 1 + \varepsilon > 1.$$

We will construct a signal $f \in \mathcal{V}_e \backslash \mathcal{V}$ for which there exists an admissible Q such that $(I - QH)f \in \mathcal{V}$, hence $(I - QH)^{-1}$ is not stable. Towards this end, there exists an $f_0 \in \mathcal{V}$ and time $t_0 \in \mathcal{T}$ such that

(1) $\operatorname{supp}(f_0) = [0, t_0]$.

(2) $\|(Hf_0)|_{[0,t_0]}\| \geq (1 + \varepsilon/2)\|f_0\|$.

Similarly, since $\|HS_t\| > 1$ uniformly in time, there exists an $f_1 \in \mathcal{V}$ and time $t_1 \in \mathcal{T}$ such that

(1) $\operatorname{supp}(f_1) = [FM(f_0, t_0; H) + 1, t_1]$.

(2) $\|(Hf_1)|_{[FM(f_0,t_0,H)+1,t_1]}\| \geq (1 + \varepsilon/2)\|f_1\|$.

This sequence of f_n and t_n has the following recursive form:

(1) $\operatorname{supp}(f_n) = [FM(f_0 + \cdots + f_{n-1}, t_{n-1}; H) + 1, t_n]$.

(2) $\|Hf_n|_{[FM(f_0+\cdots+f_{n-1},t_{n-1},H)+1,t_n]}\| \geq (1 + \varepsilon/2)\|f_n\|$.

Note that via Proposition 3.1,

$$H\left(\sum_n f_n\right) = \sum_n Hf_n.$$

Since H has the UINE property, we may assume that

$$\lim_{n \to \infty} \|f_n\| = \alpha \in (0, \infty).$$

If this were not the case, the UINE property assures that $\|f_n\|$ may be selected so that

$$0 < \alpha_{\min} \leq \|f_n\| \leq \alpha_{\max} < \infty.$$

Thus, we may select a subset of the f_n such that $\|f_n\|$ is convergent.

Now let $f = \sum_{n=0}^{\infty} f_n$. Then $f \in \mathcal{V}_e \backslash \mathcal{V}$. In order to construct the destabilizing Q, let I_n denote the

ith interval

$$I_n = [FM(f_0 + \cdots + f_{n-1}, t_{n-1}; H) + 1, t_n].$$

Then

$$\|(Hf)|_{I_n}\| \ge (1 + \varepsilon/2)\, \|f|_{I_n}\|.$$

Since $\|f_n\| \to a \to 0$, for $n \ge N^*$ with N^* sufficiently large,

$$\frac{\|f_{n+1}\|}{\|f_n\|} \le 1 + \varepsilon/4.$$

The destabilizing Q is now constructed as follows. For $n < N^*$, set $Q_n = 0$. For $n \ge N^*$, set Q_n to be the LTV operator which maps $(Hf)|_{I_n}$ to $f|_{I_{n+1}}$ as in Lemma 2.1. It follows that

$$\|Q_n\| \le \frac{1 + \varepsilon/4}{1 + \varepsilon/2} < 1, \quad n \ge N^*.$$

Now set $Q = \sum_n Q_n$. With this construction, Q exhibits pure delay hence is admissible. Furthermore, the summation $Q = \sum_n Q_n$ admits a "block diagonal" representation (cf Lemma 2.1), hence $\|Q\| < 1$. Finally,

$$(I - QH)f = \sum_{n=0}^{N^*} f_n \in \mathcal{V}$$

Since $(I - QH)f \in \mathcal{V}$ and $f \in \mathcal{V}_e \backslash \mathcal{V}$, the stable invertibility of $(I - QH)$ is violated

The case $\inf_{t \in \mathcal{J}} \|HS_t\| = 1$ follows from continuity arguments as in Shamma (1991) ∎

Note that only *linear* Qs are used for destabilization. For a time-invariant H, the above construction can always lead to a linear Q which is also *periodic*.

The main idea in the proof for finite-memory it to exploit the right-distributivity property of Proposition 3.1. In fact, one could use this property alone to define an even weaker notion of pointwise fading-memory for which Theorem 4.2 still holds.

It turns out that for $\mathcal{V} = \mathcal{L}_0^\infty$ or c_0, it is possible to construct a *destabilizing* Q for fading-memory H as follows.

Theorem 4.3. Let $\mathcal{V}$ denote either $\mathcal{L}_0^\infty$ or c_0. Let $H \in \mathcal{S}_{NL}$ have pointwise fading-memory and the UINE property. Let $Q \in \mathcal{S}_{NL}$ be called "admissible" if the operator $I - QH$ is causally invertible and $\|Q\| < 1$. Then the operator $I - QH$ has a *stable* inverse for all admissible $Q \in \mathcal{S}_L$ only if

$$\inf_{t \in \mathcal{J}} \|HS_t\| \le 1.$$

Proof. For clarity of presentation, the proof is presented for time-invariant H with $\mathcal{V} = \mathcal{L}_0^\infty$. In the time-varying case, the UINE property can be exploited as in the proof of Theorem 4.2. The proof for $\mathcal{V} = c_0$ follows from similar arguments.

Since $\|H\| > 1$, there exists an $f_0 \in \mathcal{L}_0^\infty$ and $\varepsilon > 0$ such that

(1) $\mathrm{supp}(f_0) = [0, t_0]$.

(2) $\|(Hf_0)|_{[0,t_0]}\| \ge (1 + 2\varepsilon)\, \|f_0\|$.

As in the proof of Theorem 4.2, we will construct a signal $f \in \mathcal{L}^\infty \backslash \mathcal{L}_0^\infty$ and an admissible Q such that $(I - QH)f \in \mathcal{L}_0^\infty$. This signal f will take the form

$$f = S_{T_0} f_0 + S_{T_1} f_0 + S_{T_2} f_0 + S_{T_3} f_0 + \ldots,$$

where the T_i are chosen such that the effects of the previous inputs have sufficiently decayed. Towards this end, let

$$T = 0,$$
$$T_1 = FM_\varepsilon(f_0, t_0; H) + 1,$$
$$T_2 = FM_\varepsilon(f_0 + S_{T_1} f_0, T_1 + t_0; H) + 1,$$
$$T_3 = FM_\varepsilon(f_0 + S_{T_1} f_0 + S_{T_2} f_0, T_2 + t_0; H) + 1,$$
$$\ldots \text{etc.}$$

Let $I_n = [T_n, T_{n+1})$. It follows that

$$\|(Hf)|_{I_n}\| \ge (1 + \varepsilon)\, \|f|_{I_{n+1}}\|.$$

Thus, we can construct admissible Q_n via Lemma 2.1 which map $Hf|_{I_n} \mapsto f|_{I_{n+1}}$. Finally, set $Q = \sum_n Q_n$. Since the Q_n constructed via Lemma 2.1 do not "interact", it follows that $\|Q\| < 1$. Since the Q_n exhibit pure delay, $(I - QH)$ is causally invertible. Furthermore, $(I - QH)f = f_0 \in \mathcal{L}_0^\infty$ while $f \in \mathcal{L}_e^\infty \backslash \mathcal{L}_0^\infty$. Thus $(I - QH)^{-1}$ is not stable over $\mathcal{L}_0^\infty$. ∎

This proof breaks down for $\mathcal{V}$ denoting other than $\mathcal{L}_0^\infty$ or c_0. The reason is that norm of the exciting signal

$$f = S_{T_0} f_0 + S_{T_1} f_0 + S_{T_2} f_0 + S_{T_3} f_0 + \ldots,$$

does not remain bounded for $\mathcal{V}$ other than $\mathcal{L}_0^\infty$ or c_0. In fact, we have not shown that $(I - QH)^{-1}$ is an unbounded operator. Rather, it is the asymptotic stability which was violated.

At a glance, it would seem that the condition $\|H\| < 1$ is also sufficient for robust invertibility over $\mathcal{L}_0^\infty$. However, standard small gain arguments would only assure that the operator $(I - QH)^{-1}$ is stable over $\mathcal{L}^\infty$ and not necessarily $\mathcal{L}_0^\infty$. Some additional work is required to guarantee asymptotic stability.

Theorem 4.4. Let $\mathcal{V}$ denote either $\mathcal{L}_0^\infty$ or c_0. Let $G \in \mathcal{S}_{NL}$ have pointwise fading-memory and $\|G\| < 1$. Let $(I - G)$ be causally invertible with

$(I-G)^{-1}$ continuous, i.e.

$$f_n \to f_0 \Rightarrow (I-G)^{-1}f_n \to (I-G)^{-1}f_0.$$

Then $(I-G)^{-1}$ is stable over $\mathcal{V}$.

Proof. The proof is stated for $\mathcal{V}=\mathcal{L}_0^\infty$. The case where $\mathcal{V}=c_0$ follows from similar arguments.

Standard small gain arguments show that $(I-G)^{-1}$ is finite-gain stable over $\mathcal{L}^\infty$. We will show stability over $\mathcal{L}_0^\infty$ as follows. First, we will show that if $f\in\mathcal{L}_0^\infty$ has finite duration, then $(I-G)^{-1}f$ will eventually decay, and hence belongs to $\mathcal{L}_0^\infty$. The remainder of the proof follows from continuity of $(I-G)^{-1}$.

Towards this end, given any $\varepsilon>0$, let $\tilde{G}$ be a pointwise finite-memory approximant of G such that $\|G-\tilde{G}\|\le\varepsilon$ and $\|\tilde{G}\|<1$. Let $f\in\mathcal{L}_0^\infty$ have $\operatorname{supp}(f)=[0,T_0]$. Thus $g=(I-G)^{-1}f\in\mathcal{L}^\infty$. Let $T_1=FM(g,T_0;\tilde{G})+1$. Then since

$$g=f+\tilde{G}g+(G-\tilde{G})g,$$

it follows that

$$(I-P_{T_1})g=(I-P_{T_1})\tilde{G}g+(I-P_{T_1})(G-\tilde{G})g.$$

since $\tilde{G}$ has finite-memory,

$$(I-P_{T_1})\tilde{G}g=(I-P_{T_1})\tilde{G}(I-P_{T_0})g.$$

It follows that

$$\|(I-P_{T_1})g\|\le\|\tilde{G}\|\,\|(I-P_{T_0})g\|+\varepsilon\,\|g\|.$$

Now let $T_2=FM(g,T_1;\tilde{G})+1$. Then similarly

$$\begin{aligned}\|(I-P_{T_2})g\|&\le\|\tilde{G}\|\,\|(I-P_{T_1})g\|+\varepsilon\,\|g\|\\&\le\|\tilde{G}\|^2\,\|(I-P_{T_0})g\|\\&\quad+\|\tilde{G}\|\,\varepsilon\,\|g\|+\varepsilon\,\|g\|.\end{aligned}$$

Proceeding recursively with $T_{n+1}=FM(g,T_n;\tilde{G})+1$ leads to

$$\begin{aligned}\|(I-P_{T_n})g\|&\le(1+\cdots+\|\tilde{G}\|^{n-1})\varepsilon\,\|g\|\\&\quad+\|\tilde{G}\|^n\,\|(I-P_{T_0})g\|.\end{aligned}$$

Thus

$$\limsup_{n\to\infty}\|(I-P_{T_n})g\|\le\frac{1}{1-\|\tilde{G}\|}\,\varepsilon\,\|g\|.$$

Since ε is arbitrary and $\|\tilde{G}\|\to\|G\|$ as $\varepsilon\to0$, it follows that $g\in\mathcal{L}_0^\infty$.

Thus for any f of finite duration $(I-G)^{-1}f\in\mathcal{L}_0^\infty$. Since any $f\in\mathcal{L}_0^\infty$ can be approximated by a finite duration signal, it follows from continuity that $(I-G)^{-1}$ is finite-gain stable over $\mathcal{L}_0^\infty$.

5. CLOSED-LOOP FADING-MEMORY AND ROBUST STABILIZATION

In this section, we consider the block diagram of Fig. 1. Some preliminary assumptions and definitions are as follows. The plant, P, and

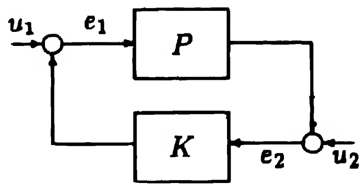


FIG. 1 Block diagram for robust stabilization.

compensator, K, are operators on $\mathcal{V}_e$. This feedback system is said to be *well-posed* (cf. Willems (1971)) if given any $(u_1,u_2)\in\mathcal{V}_e\times\mathcal{V}_e$, there exist unique $(e_1,e_2)\in\mathcal{V}_e\times\mathcal{V}_e$ which satisfy

$$e_1=u_1+Ke_2,$$
$$e_2=u_2+Pe_1.$$

such that the mapping $\Phi(P,K):\begin{pmatrix}u_1\\u_2\end{pmatrix}\mapsto\begin{pmatrix}e_1\\e_2\end{pmatrix}$ is causal. Assuming well-posedness, the compensator, K, is said to (*incrementally*) *stabilize* the plant, P, if $\Phi(P,K)$ is (incrementally) stable. The compensator K is said to *pointwise fading-memory* (*incrementally*) *stabilize* P if $\Phi(P,K)$ is (incrementally) stable and has pointwise fading-memory.

The robust invertibility conditions of Section 4 can be used to give necessary conditions for robust stability as follows (cf. Shamma (1991)). Define the following *family* of plants:

$$\mathcal{P}_{\text{add}}\overset{\text{def}}{=}\{P:P=P_0+\Delta W\},$$

where $\Delta\in\mathcal{S}_{NL}$ with $\|\Delta\|<1$, and $W\in\mathcal{S}_{NL}$. We assume that $\mathcal{P}_{\text{add}}$ is such that any causal compensator results in a well-posed feedback system for every $P\in\mathcal{P}_{\text{add}}$.

The problem of robust stabilization is under what conditions does a compensator, K, which stabilizes P_0 also stabilize every $P\in\mathcal{P}_{\text{add}}$. In case

$$\sup_{P\in\mathcal{P}_{\text{add}}}\|\Phi(P,K)\|<\infty,$$

the compensator, K, is said to *uniformly robustly stabilize* the family $\mathcal{P}_{\text{add}}$.

Theorem 5.1. Let $\mathcal{V}$ denote any one of the $\mathcal{L}_p$ or ℓ_p spaces with $p\in[1,\infty)$. Consider the plant family $\mathcal{P}_{\text{add}}$. Let the compensator, K, pointwise fading-memory stabilize P_0 such that $WK(I-P_0K)^{-1}$ satisfies the UINE property. Then K uniformly robustly stabilizes the family $\mathcal{P}_{\text{add}}$ only if

$$\inf_{t\in\mathcal{T}}\|WK(I-P_0K)^{-1}S_t\|<1.$$

In case $\mathcal{V}$ denotes $\mathcal{L}_0^\infty$ or c_0, then K robustly stabilizes the family $\mathcal{P}_{\text{add}}$ only if

$$\inf_{t\in\mathcal{T}}\|WK(I-P_0K)^{-1}S_t\|\le1.$$

Proof. The proof follows from the results of Section 4 and slight modifications of the proof of Theorem 4.1 in Shamma (1991).

A key assumption in Theorem 5.1 is that the closed-loop operator $WK(I - P_0K)^{-1}$ exhibits pointwise fading-memory. In case P is linear (possibly time-varying), it is possible to parametrize all fading-memory incrementally stabilizing compensators as follows. We will use factorization representations of P and K as in Verma (1988) to develop conditions for closed-loop fading-memory.

Definition 5.1. Let $H: \mathcal{V}_r \to \mathcal{V}_r$. Then $H = ND^{-1}$ is said to be a *right-coprime fractional representation* (*r.c.f.r.*) of H if
(1) $N, D \in \mathcal{S}_{NL}$.
(2) D is causally invertible.
(3) There exists an $F \in \mathcal{S}_{NI}$ such that

$$F\begin{pmatrix} N \\ D \end{pmatrix} = I,$$

i.e. the matrix operator $\begin{pmatrix} N \\ D \end{pmatrix}$ has an stable left inverse.

Theorem 5.2. Let $K_l = XY^{-1}$ be a linear compensator which stabilizes the linear $P = NM^{-1}$, where XY^{-1} and NM^{-1} are r.c.f.r.s with N, M, X, and Y all linear. Then all pointwise fading-memory incrementally stabilizing nonlinear compensators are parametrized by

$$K = (X + MQ)(Y + NQ)^{-1},$$

with Q incrementally stable with pointwise fading-memory.

Proof. Let $\Phi(P, K)$ be incrementally stable with pointwise fading-memory. From Verma (1988), Khargonekar and Poolla (1986), all incrementally stabilizing controllers take the given parametrized form. Since P and K_l are linear,

$$\begin{aligned}
&\begin{pmatrix} M & -(X + MQ) \\ -N & (Y + NQ) \end{pmatrix} \\
&= \begin{pmatrix} M & -X \\ -N & Y \end{pmatrix} + \begin{pmatrix} 0 & -MQ \\ 0 & NQ \end{pmatrix} \\
&= \begin{pmatrix} M & -X \\ -N & Y \end{pmatrix}\left(\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}\right. \\
&\quad \left. + \begin{pmatrix} M & -X \\ -N & Y \end{pmatrix}^{-1}\begin{pmatrix} 0 & -MQ \\ 0 & NQ \end{pmatrix}\right) \\
&= \begin{pmatrix} M & -X \\ -N & Y \end{pmatrix}\begin{pmatrix} I & -Q \\ 0 & I \end{pmatrix}.
\end{aligned}$$

Thus

$$\begin{aligned}
\Phi(P, K) &= \begin{pmatrix} I & -K \\ -P & I \end{pmatrix}^{-1} \\
&= \begin{pmatrix} M & 0 \\ 0 & Y + NQ \end{pmatrix} \\
&\quad \times \begin{pmatrix} I & Q \\ 0 & I \end{pmatrix}\begin{pmatrix} M & -X \\ -N & Y \end{pmatrix}^{-1}.
\end{aligned}$$

This implies

$$\begin{aligned}
\Phi(P, K)&\begin{pmatrix} M & -X \\ -N & Y \end{pmatrix} \\
&= \begin{pmatrix} M & 0 \\ 0 & Y + NQ \end{pmatrix}\begin{pmatrix} I & Q \\ 0 & I \end{pmatrix} \\
&= \begin{pmatrix} M & MQ \\ 0 & Y + NQ \end{pmatrix}
\end{aligned}$$

Since the left-hand-side has pointwise fading-memory, it follows that the operator MQ has pointwise fading-memory. Similarly, $Y + NQ$ has pointwise fading-memory, and hence so does NQ.

Since both MQ and NQ have pointwise fading-memory, the operator

$$\begin{pmatrix} N \\ M \end{pmatrix} Q,$$

has pointwise fading-memory. Since P was stabilized by the linear K_l, $\begin{pmatrix} N \\ M \end{pmatrix}$ has a stable linear left inverse, which implies Q has pointwise fading-memory.

To show the converse, let Q have pointwise fading-memory. That $\Phi(P, K)$ has pointwise fading-memory follows from

$$\begin{aligned}
\Phi(P, K) &= \begin{pmatrix} I & -K \\ -P & I \end{pmatrix}^{-1} \\
&= \begin{pmatrix} M & MQ \\ 0 & Y + NQ \end{pmatrix}\begin{pmatrix} M & -X \\ -N & Y \end{pmatrix}^{-1}.
\end{aligned}$$

That is, $\Phi(P, K)$ is the composition of continuous fading-memory operators.

6 CONCLUDING REMARKS

In this paper, we have investigated the fading-memory property primarily in the context of robust stability. Some possible directions are the following. One direction is the extension of these results to the case of structured dynamic uncertainty. Another direction is determining what classes of nonlinear systems have fading-memory. Finally, there is the issue of norm-computation of nonlinear systems (Nikolaou and Manousiouthakis (1989)).

REFERENCES

Boyd, S. and L. O. Chua (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Trans. on Circuits and Systems*, **CAS-32,** 1150–1161.

Chen, M. J. and C. A. Desoer (1982). Necessary and sufficient condition for robust stability of linear distributed feedback systems. *Int. J. of Control*, **35,** 255–267.

Dahleh, M. A. and Y. Ohta (1988). A necessary and sufficient condition for robust BIBO stability. *Systems and Control Letters*, **11,** 271–275.

Desoer, C. A. and M. Vidyasagar (1975). *Feedback Systems: Input–Output Properties*. Academic Press, New York.

Dorato, P. (Ed). *Robust Control*. IEEE Press, New York.

Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proceedings*, **129,** 242–250.

Doyle, J. C., and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. on Aut. Control*, **AC-26,** 4–16.

Doyle, J. C. and J. E. Wall and G. Stein (1982). Performance and robustness analysis for structured uncertainty. *Proc. of the 21st IEEE Conf. on Decision and Control*, 629–636.

Francis, B. A. (1987). *A Course in $\mathcal{H}^\infty$—Optimal Control Theory*. Springer-Verlag, New York.

Georgiou, T. T. and M. Smith (1990). Optimal robustness in the gap metric. *IEEE Trans. on Aut. Control*, **AC-35,** 673–686.

Khammash, M. and J. B. Pearson (1991). Performance robustness of discrete-time systems with structured uncertainty. *IEEE Trans. on Aut. Control*, **AC-36,** 398–412.

Khargonekar, P. P. and K. R. Poolla (1986). Uniformly optimal control of linear time-varying plants: nonlinear time-varying controllers. *Systems and Control Letters*, **5,** 303–308.

Nikolaou, M. and V. Manousiouthakis (1989). A hybrid approach to nonlinear system stability and performance. *AIChE J.*, **35,** 559–572.

Rudin, W. (1987) *Real and Complex Analysis*. McGraw-Hill, New York.

Sandberg, I.. W. (1965). An observation concerning the application of the contraction mapping fixed-point theorem and a result concerning the norm-boundedness of solutions of nonlinear functional equations. *Bell Systems Technical Journal*, **44,** 1809–1812.

Shamma, J. S. (1991). The necessity of the small gain theorem for time-varying and nonlinear systems. *IEEE Trans. on Aut. Control*, **36,** 1138–1147.

Verma, M. S. (1988). Coprime factorizational representations and stability of nonlinear feedback systems. *Int. J. of Control*, **48,** 897–918.

Willems, J. C. (1971). The Analysis of Feedback Systems. MIT Press, Cambridge, MA.

Youla, D. C., H. A. Jabr and J. J. Bongiorno, Jr (1976). Modern Wiener–Hopf design of optimal controllers: Part II. *IEEE Trans. on Aut. Control*, **AC-21,** 319–338.

Zames, G. (1966). On the input-output stability of time-varying nonlinear feedback systems. part I: conditions using concepts of loop gain, conicity, and positivity. *IEEE Trans. on Aut. Control*, **AC-11,** 228–238.

Automatica, Vol 29, No 1 pp 201 213 1993
Printed in Great Britain

0005 1098/93 $6 00 + 0 00

Bibliography on Robust Control*

P. DORATO,† R. TEMPO‡ and G. MUSCATO§

This bibliography includes a compilation of selected books and journal articles written on the subject of Robust Control during the period 1987–1991.

Key Words—Bibliography, control systems, robust control, uncertainty

Abstract—This bibliography includes a compilation of selected books and journal articles written on the subject of *Robust Control* during the period 1987–1991 Papers are further divided into theory and applications The theoretical subject matter of each article is identified by appropriate acronyms and application papers are separated into specific areas

1 INTRODUCTION

IN THE VOLUME "Robust Control" (Dorato, 1987) there is a fairly comprehensive bibliography of papers and books written on the topic of *robust control* up to the year 1986. The purpose of this compilation is to update that bibliography to cover the period 1987–1991. The only exception we have made to citations prior to 1987 is the seminal paper of Kharitonov (Kharitonov, 1978) on stability of interval polynomials. For earlier citations on this subject, The interested reader may consult the reference list in the book of Bhattacharyya (1987).

In order to keep the bibliography to a manageable size, a number of topics related to the problem of analysis and design of control systems in the presence of uncertainty have been omitted. In particular, we did not include papers and books on the following subjects: (i) Adaptive control; (ii) robust–adaptive control; (iii) fuzzy control; (iv) variable structure control; (v) stochastic control; (vi) robust estimation. For the same reason, we have also omitted conference papers, very brief papers and surveyed only the English-language literature on the subject.

* Received 7 October 1991, revised 6 April 1992, received in final form 22 June 1992 This paper was not presented at any IFAC meeting This paper was recommended for publication in revised form by Guest Editor R F Curtain

† Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico, U S A

‡ CENS-CNR, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

§ Dipartimento Elettrico, Elettronico e Sistemistico, Universitá di Catania, Viale A Doria 6, 95125 Catania, Italy

The bibliography is divided into two main parts: books and journal articles (see the list of journals in Section 3). Journal articles are further divided into theory and applications. Each paper is listed just once in the Reference section and is coded by acronyms (see the list in Section 3.1) to help the reader identify the theoretical subject matter of the article. Application papers are broken down into various areas so that readers may have easy access to applications of specific interest.

2 BOOKS

Under the book section of our reference list, we have compiled texts, lecture notes, reprint volumes, software userguides and conference proceedings on the subject of robust control published in book form during the period 1987–1991. We also include a few books printed at the beginning of 1992.

3 JOURNAL ARTICLES

Journal articles were compiled from the following journals.

ASME J. of Dynamic Systems, Measurement and Control, Automatica, IEE Proceedings—Part D, IEEE Control Systems Magazine, IEEE Transactions on Automatic Control, IEEE Transactions on Circuits and Systems, International J. of Control, International J. of Robust and Nonlinear Control, J. of Guidance, Control and Dynamics, J. of the Franklin Institute, Mathematics of Control, Signals and Systems, SIAM J. on Control and Optimization, Systems and Control Letters.

3.1. *Theory*

The theoretical subject matter of each paper is identified by appropriate acronyms (see the list below) written at the end of each citation. Many of the acronyms correspond to standard usage in the literature, for example [LQG]

for Linear-Quadratic-Gaussian. Others were created expressly for this bibliography, for example [KRF] for Kharitonov methods for rational functions. We have elected not to explicitly divide papers by the classification of structured/unstructured perturbations, although this could be a useful separation of robust control papers. Instead, we have separated our categories in terms of nominal models, theoretical approaches and performance measures. Thus, for example, state space and matrix methods [SSM] may include papers with structured or unstructured perturbations, but with state space models subsumed. Also, since the literature on H^∞ is vast, we have attempted to be more selective in this area and to include only papers directly related to robust control problems or theoretical papers of special significance.

List of acronyms

DPS—Distributed parameters/time-delay systems; DTS—Discrete-time systems/sampled-data; FDM—Frequency-domain methods; GCC—Guaranteed cost control; HHM—H^2/H^∞ methods for robust control; HINF—H^∞ methods for robust control; KHAR—Kharitonov theory and polynomial methods; KRF—Kharitonov methods for rational functions; LONE—l_1 methods for robust control; LQG—Robust linear quadratic Gaussian; LSS—Large scale systems; LTR—Loop transfer recovery; MIMO—Multi-input–multi-output; MOD—Uncertainty modelling; MSN—Mixed sensitivity methods; MSU—Mixed structured/unstructured perturbations; NLC—Nonlinear vs linear controller; NTV—Nonlinear/time-varying systems; OBD—Observer based design; QFT—Quantitative feedback theory; QSM—Quadratic stabilization methods; RCO—Robust controllability and observability; RMM—Robust model matching; RMO—Robust multiobjective design; ROD—Reduced order design; RPA—Robust pole assignment; RPM—robust performance methods; RSP—Robust stabilization problem; RST—Robust servo/tracking problems; SPR—Strictly positive real methods; SSM—State space and matrix methods; SSV—Structured singular values; SURVEY—Survey papers; TUTORIAL—Tutorial papers.

3.2. *Applications*

The application papers are separated into various areas so that the reader can easily identify the subject of interest. The appropriate acronym identifies the theory used in the work. It should be noted that most applications found in journal articles might be called more properly "feasibility studies", since they seldom deal with hardware implementation on a physical system. The application areas surveyed here include the following: DC-motors, electromechanical, engine control, flexible structures, flight control, mechanical systems, process control, robotics, thermal systems.

It should be noted that many application papers are reported at conferences, rather than published in journals and hence do not appear in this bibliography. The following conferences typically include papers on applications of robust control theory:

AIAA Conference on Guidance, Navigation and Control, AIChE Annual Meeting, American Control Conference, IEEE Conference on Decision and Control, IEEE Int. Conference on Robotics and Automation, IFAC Symposia and World Congress, Int. Symposium on Industrial Robots, Int. Symposium on Robotics Research.

4 CONCLUSIONS

The 15 year period 1972–1987 could properly be called the developmental period for what we now call *robust control theory*. Included in the theoretical developments of this period we have: H^∞ methods [HINF], guaranteed-cost control [GCC], Kharitonov methods [KHAR], loop-transfer-recovery theory [LTR], mixed-sensitivity methods [MSN], quadratic stabilization methods [QSM], robust pole assignment [RPA], robust servo/tracking methods [RST], state-space and matrix methods [SSM] and structured-singular-value methods [SSV]. However, the theory developed during this period was largely divided between structured and unstructured methods, with good analysis results for the former and good synthesis results for the latter; and between robust stability and nominal performance. This bibliography indicates a continued interest in these topics; however it also indicates a number of new directions including: robust multiobjective design methods [RMO], the most notable example being H^2/H^∞ methods [HHM]; mixed structured/unstructured perturbation methods [MSU]; l_1 methods [LONE]; robust control of nonlinear and time-varying systems [NTV]; robust control of distributed parameter systems [DPS]; robust performance methods [RPM]; and passivity methods [SPR].

A major post-1986 theoretical result was the two-Riccati solution to the H^∞ control problem (Doyle *et al.*, 1989). This work reduced the H^∞ problem to the solution of two uncoupled Riccati equations and provided an upper bound on the complexity of the H^∞ controller, equal to the

McMillan degree of the "augmented" plant. Other notable theoretical developments include new results on the mixed H^2/H^∞ problem and associated coupled Riccati equations; the design of reduced order compensators and the use of Kharitonov methods for mixed structured/unstructured problems.

In recent years, the major applications of robust control theory appear to be in the following areas: flexible structures, flight control, process control and robotics. In addition, a number of robust control software packages have been developed, including "μ Toolbox", "Robust-Control Toolbox", "Matrix$_x$" and "Program CC".

Acknowledgements—This paper was partially supported by the U.S. National Science Foundation under Grant No. INT-9016501 and by CENS-CNR of Italy. Part of this work was performed while the first author was visiting CENS-CNR. The authors wish to express their gratitude to Professors J. Kogan (University of Maryland) and Y. Z. Tsypkin (Russian Academy of Sciences) for making available their lists of references on the subject and to Mr G. Bracone for his help in compiling this bibliography. The following individuals also contributed to this paper and their inputs were appreciated: C. Abdallah, S. S. Banda, S. P. Bhattacharyya, P. P. Khargonekar, H. Kimura, M. G. Safonov and R. K. Yedavalli.

REFERENCES

Books

Bahnasavi, A. A. and M. S. Mahmoud (1989). Control of partially-known dynamical systems. *Lecture Notes in Control and Information Sciences*, Vol. 124. Springer-Verlag, Berlin.

Balas, G., J. Doyle, K. Glover, A. Packard and R. Smith (1991). *The μ Analysis and Synthesis Toolbox*. Musyn, MN.

Basar, T. and P. Bernhard (1991). *H^∞-Optimal Control and Related Minimax Design Problems*. Birkhauser, Boston.

Bhattacharyya, S. P. (1987). Robust stabilization against structured perturbations. *Lecture Notes in Control and Information Sciences*, Vol. 99. Springer-Verlag, Berlin.

Bhattacharyya, S. P. and L. H. Keel (Eds) (1991). *Control of Uncertain Dynamic Systems*. CRC Press, London.

Boyd, S. P. and C. H. Barratt (1991). *Linear Controller Design: Limits of Performance*. Prentice Hall, Englewood Cliffs, NJ.

Chiang, R. Y. and M. G. Safonov (1988). *Robust-Control Toolbox*. The Math Works, South Natick.

Conte, G., A. M. Perdon and B. Wyman (Eds) (1991). *New Trends in Systems Theory*. Birkhauser, Boston.

Curtain, R. F. (Ed.) (1987). *Modelling, Robustness and Sensitivity Reduction in Control Systems*. Springer-Verlag, Berlin.

D'Azzo, J. J. and C. H. Houpis (1988). *Linear Control System Analysis and Design: Conventional and Modern* (3rd ed.). McGraw-Hill, New York.

Dorato, P. (Ed.) (1987). *Robust Control*. IEEE Press, New York.

Dorato, P. and R. K. Yedavalli (Eds) (1990). *Recent Advances in Robust Control*. IEEE Press, New York.

Dorato, P., L. Fortuna and G. Muscato (1992). Robust control for unstructured perturbations: an introduction. *Lecture Notes in Control and Information Sciences*, Vol. 168. Springer-Verlag, Berlin.

Doyle, J. C., B. A. Francis and A. R. Tannenbaum (1992). *Feedback Control Theory*. Macmillan, New York.

Francis, B. A. (1987). A course in H^∞ control. *Lecture Notes in Control and Information Sciences*, Vol. 88. Springer-Verlag, Berlin.

Freudenberg, J. S. and D. J. Looze (1988). Frequency domain properties of scalar and multivariable feedback systems. *Lecture Notes in Control and Information Sciences*, Vol. 104. Springer-Verlag, Berlin.

Grujic, L. T., A. A. Martynyuk and M. Ribbens Pavella (1987). Large scale system stability under structural and singular perturbation. *Lecture Notes in Control and Information Sciences*, Vol. 92. Springer-Verlag, Berlin.

Gutman, S. (1990). Root clustering in parameter space. *Lecture Notes in Control and Information Sciences*, Vol. 141. Springer-Verlag, Berlin.

Habets, L. C. G. J. M. (1991). Robust stabilization in the gap-topology. *Lecture Notes in Control and Information Sciences*, Vol. 150. Springer-Verlag, Berlin.

Hammond, P. H. (Ed.) (1991). *Robust Control System Design using H^∞ and Related Methods*. Institute of Measurement and Control, London.

Hinrichsen, D. and B. Martensson (Eds) (1990). *Control of Uncertain Systems*. Birkhauser, Boston.

Integrated Systems (1990). Matrix$_x$/System Build Tutorial. Integrated Systems Inc., Santa Clara.

Joshi, S. M. (1989). Control of large flexible space structures. *Lecture Notes in Control and Information Sciences*, Vol. 131. Springer-Verlag, Berlin.

Kaashoek, M. A., J. H. van Schuppen and A. C. M. Ran (Eds) (1990). *Robust Control of Linear System and Nonlinear Control*. Birkhauser, Boston.

Lunze, J. (1988). *Robust Multivariable Feedback Control*. Prentice Hall, Englewood Cliffs, NJ.

Maciejowski, J. M. (1989). *Multivariable Feedback Design*. Addison-Wesley, Reading, MA.

McFarlane, D. C. and K. Glover (1990). Robust controller design using normalized coprime factor plant descriptions. *Lecture Notes in Control and Information Sciences*, Vol. 138. Springer-Verlag, Berlin.

Milanese, M., R. Tempo and A. Vicino (Eds) (1989). *Robustness in Identification and Control*. Plenum Press, New York.

Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall, Englewood Cliffs, NJ.

Mosca, E. and L. Pandolfi (Eds) (1990). H^∞-control theory. *Lecture Notes in Mathematics*, Vol. 1496. Springer-Verlag, Berlin.

Mustafa, D. and K. Glover (1990). Minimum entropy H^∞ control. *Lecture Notes in Control and Information Sciences*, Vol. 146. Springer-Verlag, Berlin.

Nijmeijer, H. and J. M. Schumacher (Eds) (1989). Three decades of mathematical system theory. *Lecture Notes in Control and Information Sciences*, Vol. 135. Springer-Verlag, Berlin.

Reinschke, K. J. (1988). Multivariable control. a graph theory approach. *Lecture Notes in Control and Information Sciences*, Vol. 108. Springer-Verlag, Berlin.

Subrahmanyam, M. B. (1990). Optimal control with a worst-case performance criterion and applications. *Lecture Notes in Control and Information Sciences*, Vol. 145. Springer-Verlag, Berlin.

Thompson, P. M. (1988). Program CC Reference Manual. Systems Technology Inc., Hawthorne.

Weinmann, A. (1991). *Uncertain Models and Robust Control*. Springer-Verlag, Wien.

Zinober, A. S. I. (Ed.) (1990). *Deterministic Control of Uncertain Systems*. Peter Peregrinus, London.

Journal articles: Theory

Abdallah, C., D. Docampo and R. Jordan (1991). Necessary and sufficient conditions for the stability of polynomials with linear parameter dependencies. *Int. J. of Robust and Nonlinear Control*, **1**, 69–78. [KHAR]

Abdul-Wahab, A. A. and M. A. Zohdy (1988). Generalized linear transformations on the design of robust dynamic output feedback controllers. *Int. J. Control*, **48**, 1241–1266. [RMM]

Abdul-Wahab, A. A. (1990). Robustness measure bounds for Lyapunov-type state-feedback systems. *IEE Proceedings, Part D*, **137**, 321–337. [LQG], [SSM]

Ackermann, J., H. Z. Hu and D. Kaesbauer (1990). Robustness analysis: a case study. *IEEE Trans. Aut. Control*, **AC-35**, 352–356. [KHAR]

Ackermann, J. E. and H. Z. Hu (1991). Robustness of sampled-data control systems with uncertain physical plant parameters. *Automatica*, **27**, 705–710. [DTS], [KHAR]

Ackermann, J., D. Kaesbauer and R. Muench (1991). Robust gamma-stability analysis in a plant parameter space. *Automatica*, **27**, 75–85. [KHAR]

Agathoklis, P. and V. Sreeram (1989). Lower bounds for the stability margin of linear continuous systems based on weighted performance criteria. *IEE Proceedings, Part D*, **136**, 63–67. [RSP]

Amillo, J. and F. A. Mata (1991). Robust stabilization of stystems with multiple real pole uncertainties. *IEEE Trans. Aut. Control*, **AC-36**, 749–752. [HINF], [RSP]

Anderson, B. D. O., E. I. Jury and M. Mansour (1987). On robust Hurwitz polynomials. *IEEE Trans. Aut. Control*, **AC-32**, 909–913. [KHAR]

Argoun, M. B. (1987). Stability of a Hurwitz polynomial under coefficients perturbations: necessary and sufficient conditions. *Int. J. Control*, **45**, 739–744. [FDM], [KHAR]

Argoun, M. B. (1987). Frequency domain conditions for the stability of perturbed polynomials. *IEEE Trans. Aut. Control*, **AC-32**, 913–916. [KHAR]

Badr, R. I., M. F. Hassan, J. Bernussou and A. Y Bilal (1988). Time-domain robustness and performance optimisation of linear quadratic control problems. *IEE Proceedings, Part D*, **135**, 223–231. [LQG]

Bahnasawi, A. A., A. S. Al-Fuhaid and M. S. Mahmoud (1989). Linear feedback approach to the stabilization of uncertain discrete systems. *IEE Proceedings, Part D*, **136**, 47–52. [DTS], [LQG], [SSM]

Bailey, F. N., D. Panzer and G. Gu (1988). Two algorithms for frequency domain design of robust control systems. *Int. J. Control*, **48**, 1787–1806. [KHAR], [KRF]

Balakrishnan, V., S. Boyd and S. Balemi (1991). Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems. *Int J of Robust and Nonlinear Control*, **1**, 295–317. [KHAR]

Barmish, B. R. (1989). A generalization of Kharitonov's four-polynomial concept for robust stability problems with linearly dependent coefficient perturbations *IEEE Trans Aut. Control*, **AC-34**, 157–165. [KHAR]

Barmish, B. R. and P. P. Khargonekar (1990). Robust stability of feedback control systems with uncertain parameters and unmodeled dynamics. *Mathematics of Control, Signal Systems*, **3**, 197–210. [KHAR], [MSU]

Barmish, B. R., P. P. Khargonekar, Z. C. Shi and R. Tempo (1990). Robustness margin need not be a continuous function of the problem data. *Systems and Control Letters*, **15**, 91–98. [KHAR], [SSV]

Barmish, B. R., M. Fu and S. Saleh (1988). Stability of a polytope of matrices: counterexamples. *IEEE Trans. Aut. Control*, **AC-33**, 569–572. [KHAR], [SSM]

Barmish, B. R. and Z. Shi (1989). Robust stability of perturbed systems with time delays. *Automatica*, **25**, 371–381. [DPS], [KHAR]

Barmish, B. R. and Z. Shi (1990). Robust stability of a class of polynomials with coefficients depending multilinearly on perturbations. *IEEE Trans. Aut. Control*, **AC-35**, 1040–1043. [KHAR]

Barmish, B. R. and R. Tempo (1990). The robust root locus. *Automatica*, **26**, 283–292. [KRF]

Barmish, B. R. and R. Tempo (1991). On the spectral set for a family of polynomials. *IEEE Trans. Aut. Control*, **AC-36**, 111–115. [KHAR]

Bartlett, A. C. and C. V. Hollot (1988). A necessary and sufficient condition for Schur invariance and generalized stability of polytopes of polynomials. *IEEE Trans. Aut. Control*, **AC-33**, 575–578. [DTS], [KHAR]

Bartlett, A. C., C. V. Hollot and Huang Lin (1987). Root locations of an entire polytope of polynomials: it suffices to check the edges. *Mathematics of Control, Signal Systems*, **1**, 61–71. [KHAR]

Basar, T. (1991). A dynamic games approach to controller design: disturbance rejection in discrete-time. *IEEE Trans. Aut. Control*, **AC-36**, 936–952. [DTS], [HINF]

Basile, G. and G. Marro (1987). On the robust controlled invariant. *Systems and Control Letters*, **9**, 191–195. [RCO], [SSM]

Basu, S. (1989). On the multidimensional generalization of robustness of scattering Hurwitz property of complex polynomials. *IEEE Trans. Circ. and Systems*, **CAS-36**, 1159–1167. [KHAR]

Beale, S. and B. Shafai (1989). Robust control system design with a proportional integral observer. *Int. J. Control*, **50**, 97–111. [OBD]

Bentsman, J., J. Fakhfakh and B. Lehman (1989). Vibrational stabilization of linear time delay systems and is robustness with respect to the delay size. *Systems and Control Letters*, **12**, 267–272. [DPS], [RSP]

Berger, C. S. (1989). Robust control of discrete systems. *IEE Proceedings, Part D*, **136**, 165–170. [DTS], FDM]

Bernstein, D. S. (1987). Robust static and dynamic output-feedback stabilization: deterministic and stochastic perspectives. *IEEE Trans. Aut. Control*, **AC-32**, 1076–1084. [SSM], [RSP]

Bernstein, D. S. (1989). Robust stability and performance via fixed-order dynamic compensation. *SIAM J. on Control and Optimization*, **27**, 389–406. [RPM], [SSM]

Bernstein, D. S. and W. M. Haddad (1988). The optimal projection equations with Petersen-Hollot bounds: robust stability and performance via fixed-order dynamic compensation for systems with structured real-valued parameter uncertainty. *IEEE Trans. Aut. Control*, **AC-33**, 578–582. [RPM], [SSM]

Bernstein, D. S. and W Haddad (1989). LQG control with an H^{∞} performance bound: a Riccati equation approach. *IEEE Trans. Aut. Control*, **AC-34**, 293–305. [HHM], [LQG]

Bernstein, D S. and W M Haddad (1989) Robust decentralized static output feedback. *Systems and Control Letters*, **12**, 309–318. [LSS]

Bernstein, D. S. and W. M. Haddad (1989). Robust stability and performance analysis for linear dynamic systems. *IEEE Trans Aut. Control*, **AC-34**, 751–758. [RPM]

Bernstein, D. S. and W. M. Haddad (1990). Robust stability and performance via fixed-order dynamic compensation with guaranteed cost bounds. *Mathematics of Control, Signal Systems*, **3**, 139–163. [RPM]

Bernstein, D. S. and C. V. Hollot (1989). Robust stability for sampled data control systems. *Systems and Control Letters*, **13**, 217–226. [DTS]

Bernussou, J., P. L. D Peres and J. C. Geromel (1989). A linear programming oriented procedure for quadratic stabilization of uncertain systems *Systems and Control Letters*, **13**, 65–72 [QSM]

Biernacki, R. M., H Hwang and S. P. Bhattacharyya (1987). Robust stability with structured real parameter perturbations. *IEEE Trans Aut. Control*, **AC-32**, 495–506. [KRF]

Birdwell, J. D. and A. J. Laub (1987). Balanced singular values for LQG/LTR design. *Int. J. Control*, **45**, 939–950. [LQG], [LTR]

Bistritz, Y (1988) Stability criterion for continuous-time system polynomials with uncertain complex coefficients. *IEEE Trans. Circ. and Systems*, **CAS-35**, 442–448. [KHAR]

Blanchini, F. (1990). Feedback control for linear time-invariant systems with state and control bounds in the presence of disturbances. *IEEE Trans Aut. Control*, **AC-35**, 1231–1234. [QSM]

Bose, N. K., E. I. Jury and E. Zeheb (1988). On robust Hurwitz and Schur polynomials. *IEEE Trans. Aut. Control*, **AC-33**, 1166–1168. [DTS], [KHAR]

Bose, N. K. and K. D. Kim (1989). Stability of a complex polynomial set with coefficients in a diamond and generalizations. *IEEE Trans. Circ. and Systems*, **CAS-36**, 1168–1174. [KHAR]

Bourles, H., Y. Joannic and O. Mercier (1990). Rho-stability and robustness: discrete-time case. *Int. J Control*, **52**, 1217–1239. [DTS]

Bourles H (1987) Alfa stability and robustness of large-scale interconnected systems *Int J Control* **45**, 2221–2232 [LSS]

Boyd S V Balakrishnan, C Barrat N Krais hi X Li D Mayer and S Norman (1988) A new method and associated architectures for linear controllers *IEEE Trans Aut Control* **AC-33**, 268–283 [RMO]

Boyd, S and Q Yang (1989) Structured and simultaneous Lyapunov functions for system stability problems *Int J Control* **49**, 2215–2240 [QSM] [SPR] [SSM]

Boyd, S V Balakrishnan and P Kabamba (1989) A bisection method for computing the H^∞ norm of a transfer matrix and related problems *Mathematics of Control Signal and Systems* **2**, 207–219 [HINF]

Byers R and S G Nash (1989) Approaches to robust pole assignment *Int J Control* **49**, 97–117 [RPA]

Chang B C O Ekdal H H Yeh and S S Banda (1991) Computation of the real structured singular value via polytopic polynomials *J of Guidance Control and Dynamics* **14**, 140 147 [SSV]

Chang J (1991) Robust finite settling time stabilization of discrete time system with structured uncertainty *Systems and Control Letters* **16**, 49–58 [DTS] [RSP]

Chang, K C and H C Lu (1990) Robust stabilization for multivariable systems with constrained control frequency domain approach *Mathematics of Control Signal Systems* **3**, 81–96 [FDM] [MIMO] [RSP]

Chapellat H and S P Bhattacharyya (1989) A generalization of Kharitonov s theorem robust stability of interval plants *IEEE Trans Aut Control* **AC-34**, 306–311 [KRF]

Chapellat H S P Bhattacharyya and M Dahleh (1990) Robust stability of a family of a disc polynomials *Int J Control* **51**, 1353–1362 [KHAR]

Chapellat H M Dahleh and S P Bhattacharyya (1990) Robust stability under structured and unstructured perturbations *IEEE Trans Aut Control* **AC-35**, 1100–1108 [KRF] [MSU]

Chapellat H M Dahleh and S P Bhattacharyya (1991) On robust non linear stability of interval control systems *IEEE Trans Aut Control* **AC-36**, 59 67 [KRF] [NTV]

Charlet B (1987) Stability and robustness for nonlinear systems decoupled and linearized by feedback *Systems and Control Letters* **8**, 367–374 [NTV]

Chen B M A Saberi and P Sannuti (1991) A new stable compensator design for exact and approximate loop transfer recovery *Automatica* **27**, 257 277 [LTR]

Chen B S and C L Lin (1989) Robust controllers for multivariable model matching systems *Int J Control* **50**, 1717–1730 [MIMO] [RMM]

Chen B S and W J Wang (1990) Robust stabilization of nonlinearly perturbed large scale systems by decentralized observer controller compensators *Automatica* **28**, 1035 1041 [LSS]

Chen B S and C C Wong (1987) Robust linear controller design time domain approach *IEEE Trans Aut Control* **AC-32**, 161–164 [MIMO] [SSM]

Chen B S and W S You (1987) Robust stabilization in an observer controller feedback system under nonlinear time varying perturbations or unmodeled dynamics *IEEE Trans Aut Control* **AC-32**, 1131 1135 [NTV] [RSP]

Chen C L and N Munro (1991) Calculation of the largest generalized stability hypersphere in the robust stability problem for the maximum settling time and minimum damping ratio cases *IEEE Trans Aut Control* **AC-36**, 756–759 [KHAR] [KRF]

Chen G and R J P de Figueiredo (1989) On robust stabilization of nonlinear control systems *Systems and Control Letters* **12**, 373–379 [NTV]

Chen, K H and W L Chen (1990) Robust controller analysis of linear multiloop feedback system using approximate models *Int J Control* **51**, 353 363 [MIMO]

Chen Y H (1987) Robust output feedback controller direct design *Int J Control* **46**, 1083–1091 [MIMO] [OBD] [SSM]

Chen, Y H (1987) Robust output feedback controller indirect design *Int J Control* **46**, 1093–1103 [MIMO] [OBD] [SSM]

Chen Y H (1988) Decentralized robust control system design for large scale uncertain systems *Int J Control* **47**, 1195 1205 [LSS]

Chen Y H (1988) Design of robust controllers for uncertain dynamical systems *IEEE Trans Aut Control* **AC-33**, 487 491 [NTV]

Chen Y H (1990) Robust control system design non adaptive versus adaptive *Int J Control* **51**, 1457 1477 [SSM]

Chen Y H and G Leitmann (1987) Robustness of uncertain systems in the absence of matching assumptions *Int J Control* **45**, 1527 1542 [SSM]

Cheres E S Gutman and Z J Palmor (1989) Stabilization of uncertain dynamic systems including state delay *IEEE Trans Aut Control* **AC-34**, 1199 1203 [DPS] [RSP]

Cheung M F and S Yurkovich (1988) On the robustness of MEOP design versus asymptotic LQG synthesis *IEEE Trans Aut Control* **AC-33**, 1061 1065 [LQG]

Chien C M and B C Wang (1990) Design technique for a SISO system with uncertain nonminimum phase plant *IEE Proceedings Part D* **137**, 381 384 [FDM]

Chou J H (1990) Stability robustness of linear state space models with structured perturbations *Systems and Control Letters* **15**, 207 210 [SSM]

Chou J H (1991) Pole assignment robustness in a specified disk *Systems and Control Letters* **16**, 41 44 [RPA]

Chou J H and I R Horng (1989) On sufficient conditions for the stability of interval matrices *J of the Franklin Institute* **326**, 19 26 [SSM]

Choura S and S Jayasuriya (1990) Robust finite time settling control of distributed parameter systems *Int J Control* **52**, 1425 1453 [DPS]

Cobb J D (1988) Toward a theory of robust compensation for systems with unknown parasitics *IEEE Trans Aut Control* **AC-33**, 1130 1138 [MOD]

Cobb J D (1988) Linear compensator design based exclusively on input output information are never robust with respect to unmodeled dynamics *IEEE Trans Aut Control* **AC-33**, 559 563 [SSM]

Curtain R F (1990) Robust stabilizability of normalized coprime factors the infinite dimensional case *Int J Control* **51**, 1173 1190 [DPS]

Dahleh M A and Y Ohta (1988) A necessary and sufficient condition for robust BIBO stability *Systems and Control Letters* **11**, 271 275 [HINF] [LONL]

Dahleh M A and J B Pearson (1988) Optimal rejection of persistent disturbances robust stability and mixed sensitivity minimization *IEEE Trans Aut Control* **AC-33**, 722 731 [MSN]

Dasgupta S (1988) Kharitonov s theorem revisited *Systems and Control Letters* **11**, 381 384 [KHAR]

Dasgupta S P J Parker B D Anderson F J Kraus and M Mansour (1991) Frequency domain conditions for the robust stability of linear and nonlinear dynamical systems *IEEE Trans Circ Systems* **CAS-38**, 389 397 [FDM]

Datko R (1988) Not all feedback stabilized hyperbolic systems are robust with respect to small time delays in their feedback *SIAM J on Control and Optimization* **26**, 697 713 [DPS]

Davison E J and B M Scherzinger (1987) Perfect control of the robust servomechanism problem *IEEE Trans Aut Control* **AC-32**, 689–702 [RST]

de Gaston R R E and M G Safonov (1988) Exact calculation of the multiloop stability margin *IEEE Trans Aut Control* **AC-33**, 156–191 [KHAR]

Dickman A (1987) On the robustness of multivariable linear feedback systems in state space representation *IEEE Trans Aut Control* **AC-32**, 407–410 [MIMO] [SSM]

Djaferis T E (1987) Robust pole assignment for systems with parameters *Int J Control* **46**, 1461–1475 [RPA]

Djaferis T E (1987) Achieving robust stability by sensing additional system outputs *IEEE Trans Aut Control*, **AC-32**, 228 231 [MIMO] [SSM]

Djaferis T E (1991) Finite frequency tests for robust performance of systems with real parameter uncertainty

IEEE Trans. Aut. Control, **AC-36,** 1477–1482. [KHAR], [RPM]

Djaferis, T. E. (1991). Representations of controllers that achieve robust performance for systems with real parameter uncertainty. *Systems and Control Letters,* **16,** 329–339. [KHAR], [RPM]

Djaferis, T. E. (1991). To stabilize a k real parameter affine family of plants it suffices to simultaneously stabilize 4^k polynomials. *Systems and Control Letters,* **16,** 187–193. [KRF], [RSP]

Djaferis, T. E. (1991). Simple robust stability tests for polynomials with real parameter uncertainty. *Int. J. Control,* **53,** 907–927. [KHAR]

Djaferis, T. E. and C. V. Hollot (1989). A Routh-like test for the stability of families of polynomials with linear uncertainty. *Systems and Control Letters,* **13,** 23–29. [KHAR]

Djaferis, T. E. and C. V. Hollot (1989). The stability of a family of polynomials can be deduced from a finite number $O(k^3)$ of frequency checks. *IEEE Trans. Aut. Control,* **AC-34,** 982–986. [KHAR]

Dolezal, V. (1991). Sensitivity and robust stability of general input–output systems. *IEEE Trans. Aut. Control,* **AC-36,** 539–550. [NTV]

Doraiswami, R. (1987). Measures of overshoot, speed of response and robustness. *Automatica,* **23,** 771–774. [DTS], [SSM]

Dorato, P., H. B. Park and Y. Li (1989). An algorithm for interpolation with units in H^∞, with applications to feedback stabilization. *Automatica,* **25,** 427–430. [HINF], [RSP]

Dorato, P. (1989). A historical review of robust control. *IEEE Control Systems Magazine,* Apr. 1989, 44–47. [HINF], [SURVEY]

Dorato, P. and Y. Li (1991). U-parameter design of robust SISO systems. *IEEE Trans. Aut. Control,* **AC-36,** 971–975. [HINF], [RMO]

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State-space solutions to standard H^2 and H^∞ control problems. *IEEE Trans. Aut. Control,* **AC-34,** 831–847. [HINF], [SSM]

Esfandiari, F. and H. K. Khalil (1989). On the robustness of the sampled-data control to unmodeled high-frequency dynamics. *IEEE Trans. Aut. Control,* **AC-34,** 900–903. [DTS]

Eslami, M. (1990). Robust nonlinear dynamic systems: the old question with a new answer. *Systems and Control Letters,* **14,** 445–451. [NTV]

Eslami, M. (1990). Computer-aided determination of stability robustness measure of linear discrete-time systems. *Automatica,* **26,** 623–627. [DTS]

Fan, M. K. H., A. L. Tits and J. C. Doyle (1991). Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics. *IEEE Trans. Aut. Control,* **AC-36,** 25–38. [MSU], [SSV]

Faxun, L. (1989). Simple criteria for stability interval polynomials. *Int. J. Control,* **50,** 339–347. [KHAR]

Ferreira, P. M. G. (1989). DDP and model following problem revisited. *Systems and Control Letters,* **12,** 129–132. [MIMO], [RST]

Foo, Y. K. and I. Postlethwaite (1988). Extensions of the small-μ test for robust stability. *IEEE Trans. Aut. Control,* **AC-33,** 172–176. [SSV]

Foo, Y. K. and Y C. Soh (1989). Root clustering of interval polynomials in the left-sector. *Systems and Control Letters,* **13,** 239–245. [KHAR]

Foo, Y. K. and Y. C. Soh (1990). A generalization of strong Kharitonov's theorem to polytopes of polynomials. *IEEE Trans. Aut. Control,* **AC-35,** 936–939. [KHAR]

Francis, B. A. and J. C. Doyle (1987). Linear control theory with an H^∞ optimality criterion. *SIAM J. on Control and Optimization,* **25,** 815–820. [HINF], [SURVEY]

Fruchter, G. (1991). Generalized zero sets location and absolute robust stabilization of continuous nonlinear control systems. *Automatica,* **27,** 501–512. [NTV], [RSP]

Fruchter, G., U. Srebro and E. Zeheb (1987). On several variable zero sets and applications to MIMO robust feedback stabilization. *IEEE Trans. Circ. and Systems,* **CAS-34,** 1208–1220. [MIMO], [RSP]

Fu, M. (1990). Computing the frequency response of linear systems with parametric perturbation. *Systems and Control Letters,* **15,** 45–52. [FDM], [KRF]

Fu, M. (1991). A class of weak Kharitonov regions for robust stability of linear uncertain systems. *IEEE Trans. Aut. Control,* **AC-36,** 975–978. [KHAR]

Fu, M. and B. R. Barmish (1988). Maximal unidirectional perturbation bounds for stability of polynomials and matrices. *Systems and Control Letters,* **11,** 173–179. [KHAR]

Fu, M., A. W. Olbrot and M. P. Polis (1989). Robust stability for time-delay systems: the edge theorem and graphical tests. *IEEE Trans. Aut. Control,* **AC-34,** 813–820. [DPS], [KHAR]

Fu, M., A. W. Olbrot and M. P. Polis (1989). Introduction to the parametric approach to robust control. *IEEE Control Systems Magazine,* Aug. 1989, 7–11. [KHAR], [TUTORIAL]

Furuta, K. (1987). Alternative robust servo-control system and its digital control. *Int. J. Control,* **45,** 183–194. [DTS], [RSP], [RST]

Garofalo, F. and G. Leitmann (1989). Guaranteeing ultimate boundedness and exponential rate of convergence for a class of nominally linear uncertain systems. *ASME J. of Dynamic Systems, Measurement and Control,* **111,** 584–588. [NTV], [QSM]

Genesio, R. and A. Tesi (1988). Results on the stability robustness of systems with state space perturbations *Systems and Control Letters,* **11,** 39–46. [SSM]

Georgiou, T. T., A. M. Pascoal and P. P. Khargonekar (1987). On the robust stabilizability of uncertain linear time-invariant plants using nonlinear time-varying controllers. *Automatica,* **23,** 617–624. [NLC]

Geromel, J. C. and J. J. DaCruz (1987). On the robustness of optimal regulators for nonlinear discrete-time systems. *IEEE Trans. Aut. Control,* **AC-32,** 703–710. [DTS], [NTV]

Geromel, J. C., P. L. D. Peres and J. Bernussou (1991). On a convex parameter space method for a linear control design of uncertain systems. *SIAM J. on Control and Optimization,* **29,** 381–402. [SSM]

Gilbert, E. G. and I. J. Ha (1987). Robust tracking in nonlinear systems. *IEEE Trans. Aut. Control,* **AC-32,** 763–771. [NTV], [RST]

Glad, S. T. (1987). Robustness of nonlinear state feedback: a survey. *Automatica,* **23,** 425–435. [NTV], [SURVEY]

Glover, K. and J. C. Doyle (1988). State space formulae for all stabilizing controllers that satisfy an H^∞-norm bound and relations to risk sensitivity. *Systems and Control Letters,* **11,** 167–172. [HINF]

Glover, K and D. McFarlane (1989). Robust stabilization of normalized coprime factor plant descriptions with H^∞-bounded uncertainty. *IEEE Trans. Aut. Control,* **AC-34,** 821–830. [HINF], [RSP]

Goodall, D. P. and E. P. Ryan (1988). Feedback controlled differential inclusions and stabilization of uncertain dynamical systems. *SIAM J. on Control and Optimization,* **26,** 1431–1447. [NTV]

Grasselli, O. M. and S. Longhi (1991). Robust output regulation under uncertainties of physical parameters. *Systems and Control Letters,* **16,** 33–40. [MIMO], [SSM]

Grimble, M. J. (1989). Extensions to H^∞ multivariable robust controllers and the relationship to LQG design problems. *Int. J. Control,* **50,** 309–338. [HINF]

Grujic, L. T. and D. B. Petkovski (1987). Robust absolutely stable Lurie systems. *Int. J. Control,* **46,** 357–368. [NTV]

Gu, K., Y. H. Chen, M. A. Zohdy and N. K. Loh (1991). Quadratic stabilizability of uncertain systems: a two level optimization setup. *Automatica,* **27,** 161–165. [QSM]

Gu, K., M. A. Zohdy and N. K. Loh (1990). Necessary and sufficient conditions of quadratic stability of uncertain linear systems. *IEEE Trans. Aut. Control,* **AC-35,** 601–604. [QSM]

Guerin, J. P. and C. Z. Xu (1987). H^∞ robust stabilizability. *Systems and Control Letters,* **9,** 289–293. [HINF]

Ha, I. J. and E. G. Gilbert (1987). Robust tracking in nonlinear systems. *IEEE Trans. Aut. Control*, **AC-32**, 763–771. [NTV], [RST]

Haddad, W. M. and D. S. Bernstein (1988). Unified optimal projections equations for simultaneous reduced-order, robust modelling, estimation and control. *Int. J Control*, **47**, 1117–1132. [SSM]

Hammer, J. (1989). Robust stabilization of non-linear systems. *Int. J. Control*, **49**, 629–653. [NTV], [RSP]

Hammer, J. (1991). Preservation of stability under system perturbations in nonlinear control: a simple characterization. *Systems and Control Letters*, **16**, 435–445. [NTV]

Hannsgen, K B., Y. Renardy and R. L. Wheeler (1988) Effectiveness and robustness with respect to time delays of boundary feedback stabilization in one-dimensional viscoelasticity. *SIAM J. on Control and Optimization*, **26**, 1200–1234. [DPS]

Hara, S. and T. Sugie (1988). Independent parametrization of two-degree-of-freedom compensators in general robust tracking systems *IEEE Trans. Aut. Control*, **AC-33**, 59–67. [RST]

Hertz, D., E. I. Jury and E. Zeheb (1987). Root exclusion from complex polydomains and some of its applications *Automatica*, **23**, 399–404. [KHAR]

Hinrichsen, D., B. Kelb and A. Linnemann (1989). An algorithm for the computation of the structured complex stability radius. *Automatica*, **25**, 771–775. [KHAR], [SSM]

Hinrichsen, D. and A. J. Pritchard (1990). A note on some differences between real and complex stability radii. *Systems and Control Letters*, **14**, 401–408. [SSM]

Hollot, C. V. and F. Yang (1990) Robust stabilization of interval plants using lead or lag compensators. *Systems and Control Letters*, **14**, 9–12. [KRF], [RSP]

Hollot, C. V., D. P. Looze and A. C. Bartlett (1990). Parametric uncertainty and unmodeled dynamics : analysis via parameter space methods. *Automatica*, **26**, 269–282. [KRF], [MSU]

Horowitz, I. (1991). Survey of quantitative feedback theory (QFT). *Int J. Control*, **53**, 255–291 [QFT], [SURVEY]

Howitt, G. D. and R. Luus (1991). Simultaneous stabilization of linear single-input systems by linear state feedback control. *Int. J. Control*, **54**, 1015–1030. [RSP], [SSM]

Hsiao, F. H. and B. S. Chen (1988). LQG optimal and robust controller design for nonlinear stochastic systems. multivariable case. *J. of the Franklin Institute*, **325**, 189–206. [LQG], [MIMO], [NTV]

Hsiao, F. H. and J. G. Hsieh (1991). Robust LQG optimal controller synthesis against noise spectral uncertainties and nonlinear time-varying unmodeled dynamics. *Systems and Control Letters*, **16**, 167–178. [LQG], [NTV]

Hyland, D. C. and D. S. Bernstein (1987). The majorant Lyapunov equation: a nonnegative matrix equation for robust stability and performance of large scale systems. *IEEE Trans. Aut. Control*, **AC-32**, 1005–1013. [LSS], [RPM], [SSM]

Hyland, D. C. and E. G. Collins (1989) An M-matrix and majorant approach to robust stability and performance analysis for systems with structured uncertainty. *IEEE Trans. Aut. Control*, **AC-34**, 699–710. [RPM], [SSM]

Hyland, D. C. and E. G. Collins (1991). Some majorant robustness results for discrete-time systems. *Automatica*, **27**, 167–172. [DTS], [SSM]

Iftar, A. and U. Ozguner (1989). An optimal control approach to the decentralized robust servomechanism problem. *IEEE Trans. Aut. Control*, **AC-34**, 1268–1271. [LSS], [RST]

Iftar, A. and U. Ozguner (1991). Modeling of uncertain dynamics for robust controller design in state space. *Automatica*, **27**, 141–146. [MOD]

Ioinovich, A. (1987). Application of the quantitative synthesis of feedback systems with uncertain non-linear plants. *Int. J. Control*, **45**, 579–587. [NTV], [QFT]

Ionescu, D. and J. Y. Lin (1990). Optimal robust stabilization compensator design: a deterministic approach. *Int. J. Control*, **51**, 585–599. [RSP]

Ishihara, T. (1988). Robust stability bounds for a class of discrete-time regulators with computation delays. *Automatica*, **24**, 697–700. [DTS]

Jabbari, F and W. E Schmitendorf (1990). A noniterative method for the design of linear robust controllers. *IEEE Trans. Aut. Control*. **AC-35**, 954–957. [SSM]

Jabbari, F amd W. E. Schmitendorf (1991). Robust linear controllers using observers. *IEEE Trans. Aut. Control*, **AC-36**, 1509–1514. [OBD]

Jacobs, O. L. R. (1989). Cost of uncertainty about controlled objects *IEE Proceedings, Part D*, **136**, 177–187. [LQG]

Jayasuriya, S. (1987). Robust tracking for a class of uncertain linear systems. *Int. J Control*, **45**, 875–892. [RST]

Jayasuriya, S. and C. D. Kee (1988). Circle-type criterion for synthesis of robust tracking controllers *Int. J Control* **48**, 865–886. [RST]

Jiang, C. (1991) New method of robust pole assignment by output feedback. *IEE Proceedings, Part D*, **138**, 172–176. [RPA]

Johnston, R. D. and G. W. Barton (1987). Control system performance robustness. *Int. J. Control*, **45**, 641–648. [RPM]

Joshi, S M. (1991). Generalized proportional-plus-derivative compensators for a class of uncertain plant. *J. of Guidance, Control and Dynamics*, **14**, 677–680. [SPR]

Juang, Y T (1989) Eigenvalue assignment robustness: the analysis for autonomous system matrices *Int. J. Control*, **49**, 1787–1797 [RPA], [SSM]

Juang, Y T., Z. C. Hong and Y. T. Wang (1989). Lyapunov approach to robust pole-assignment analysis. *Int. J. Control*, **49**, 921–927. [RPA], [SSM]

Juang, Y T., T S. Kuo and C. F. Hsu (1987). Stability robustness analysis of digital control systems in state-space models. *Int J. Control*, **46**, 1547–1556. [DTS], [SSM]

Juang, Y. T., T. S. Kuo, C. F. Hsu and S. D. Wang (1987). Root-locus approach to the stability analysis of interval matrices. *Int. J. Control*, **46**, 817–822. [SSM]

Kang, H. I., B. R. Barmish, R. Tempo and C. V. Hollot (1991) Counterexample and correction to a recent result on robust stability of a diamond of complex polynomials. *IEEE Trans. Circ. and Systems*, **CAS-38**, 1370–1373. [KHAR]

Karkanias, N. and X Y. Shan (1991) Root regions of bounded coefficients polynomials. *Int. J Control*, **53**, 929–949. [KHAR]

Katayama, T. and S. Sasaki (1987). Robust stability of linear quadratic state feedback regulator under system uncertainty. *Int J. Control*, **45**, 391–405. [LQG], [SSM]

Katbab, A. and E I. Jury (1990). Robust Schur-stability of control systems with interval plants *Int. J. Control*, **51**, 1341–1352. [DTS], [KRF]

Katbab, A and E I. Jury (1991). Generalization and comparison of two recent frequency-domain stability robustness results. *Int J. Control*, **53**, 463–475. [FDM], [KHAR]

Katbab, J. and E. I. Jury (1990). Robust Schur stability of a complex-coefficients polynomial set with coefficients in a diamond. *J. of the Franklin Institute*, **327**, 687–698. [DTS], [KHAR]

Katbab, J., K. Kraus and E I. Jury (1990). Some Schur-stability criteria for uncertain systems with complex coefficients *IEEE Trans. Circ. systems*, **CAS-37**, 1171–1176. [DTS], [KHAR]

Kautsky, J. and N. K. Nichols (1990). Robust pole assignment in systems subject to structured perturbations. *Systems and Control Letters*, **15**, 373–380. [RPA]

Keel, L. H., S. P. Bhattacharyya and J. W. Howze (1988). Robust control with structured perturbations. *IEEE Trans. Aut. Control*, **AC-33**, 68–78. [KHAR], [KRF]

Khammash, M. and J. B. Pearson (1990). Robust disturbance rejection in l^1 optimal control systems. *Systems and Control Letters*, **14**, 93–101. [LONE]

Khammash, M. and J. B Pearson (1991). Performance robustness of discrete-time systems with structured uncertainty. *IEEE Trans. Aut. Control*, **AC-36**, 398–412. [DTS], [LONE], [RPM]

Khargonekar, P P., T. T. Georgiou and A. M. Pascoal (1987) On the robust stabilizability of linear time-

invariant plants with unstructured uncertainty. *IEEE Trans. Aut. Control*, **AC-32**, 201–207. [HINF], [NLC], [RSP]

Khargonekar, P. P., A. M. Pascoal and R. Ravi (1988). Strong, simultaneous and reliable stabilization of linear time-varying plants. *IEEE Trans. Aut. Control*, **AC-33**, 1158–1161. [NTV], [RSP]

Khargonekar, P. P., I. R. Petersen and K. Zhou (1990). Robust stabilization of uncertain linear systems: quadratic stabilizability and H^∞ control theory. *IEEE Trans. Aut. Control*, **AC-35**, 356–361. [HINF], [QSM], [RSP]

Khargonekar, P. P. and M. A. Rotea (1991). Mixed H^2/H^∞ control: a convex optimization approach *IEEE Trans. Aut. Control*, **AC-36**, 824–837. [HHM]

Kharitonov, V. L. (1978). Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differential'nyfe Uraveniya*, **14**, 2086–2088. [KHAR]

Khorasani, K. (1989). Robust stabilization of non-linear systems with unmodeled dynamics. *Int. J. Control*, **50**, 827–844. [NTV], [RSP]

Kim, J. Y. and M. S. Ko (1991). Quadratic robust tracking control—A polynomial matrix approach. *Int. J. of Robust and Nonlinear Control*, **1**, 215–228. [RST]

Kim, S. W., W. H. Kwon and J. H. Lee (1991). Allowable parameter variations and robustness recovery in LQG regulators. *Int. J. of Robust and Nonlinear Control*, **1**, 33–42. [LQG], [LTR]

Kimura, H. (1987). Directional interpolation approach to H^∞-optimization and robust stabilization. *IEEE Trans. Aut. Control*, **AC-32**, 1085–1093. [HINF]

Kimura, H. (1989). Conjugation, interpolation and model-matching in H^∞. *Int. J. Control*, **49**, 269–307. [HINF]

Kokame, H. and T. Mori (1991). A Kharitonov-like theorem for interval polynomial matrices. *Systems and Control Letters*, **16**, 107–116. [KHAR]

Kolla, S. P., R. K. Yedavalli and J. B. Farison (1989). Robust stability bounds on time-varying perturbations for state-space models of linear discrete-time systems. *Int. J. Control*, **50**, 151–159. [DTS], [SSM]

Kosmidou, O I. (1990). Robust stability and performance of systems with structured and bounded uncertainties: an extension of the guaranteed cost control approach. *Int. J. Control*, **52**, 627–640. [GCC], [RMO], [RPM]

Kosmidou, O. I. and P. Bertrand (1987). Robust-controller design for systems with large parameter variations. *Int. J Control*, **45**, 927–938. [LTR], [RPM]

Kouvaritakis, B., J. A. Rossiter and J. Wang (1991). Generalized Nyquist bands for structured and highly structured uncertainty. *Int. J. Control*, **53**, 1295–1309. [FDM]

Kouvaritakis, B. and M. S. Trimboli (1989). Robust multivariable feedback design. *Int. J. Control*, **50**, 1327–1377. [MIMO]

Kraus, F. J., B. D. O. Anderson, E. I. Jury and M. Mansour (1988). On the robustness of low-order Schur polynomials *IEEE Trans. Circ. and Systems*, **CAS-35**, 570–577. [DTS], [KHAR]

Kraus, F. J., B. D. O. Anderson and M. Mansour (1988). Robust Schur polynomial stability and Kharitonov's theorem. *Int. J. Control*, **47**, 1213–1225. [DTS], [KHAR]

Kraus, F. J., B. D. O. Anderson and M. Mansour (1989). Robust stability of polynomials with multilinear parameter dependence. *Int. J. Control*, **50**, 1745–1762. [KHAR]

Kraus, F. J. and W. Truol (1991). Robust stability of control systems with polytopical uncertainty: a Nyquist approach. *Int. J. Control*, **53**, 967–983. [KRF]

Kravaris, C. and S. Palanki (1988) A Lyapunov approach for robust nonlinear state feedback synthesis. *IEEE Trans. Aut. Control*, **AC-33**, 1188–1191. [NLC], [SSM]

Laughlin, D. L., D. E. Rivera and M. Morari (1987). Smith predictor design for robust performance *Int. J. Control*, **46**, 477–504. [DPS], [RPM]

Leal, M. A. and J. S. Gibson (1990). A first-order Lyapunov robustness method for linear systems with uncertain parameters. *IEEE Trans. Aut. Control*, **AC-35**, 1068–1071. [SSM]

Lee, T. T. and M. S. Chen (1987). Robustness recovery of LQG-based multivariable control designs. *Int. J. Control*, **45**, 1131–1136. [LQG], [MIMO]

Leithead, W. E. and J. O'Reilly (1991). Uncertain SISO systems with fixed stable minimum-phase controllers: relationship of closed-loop systems to plant RHP poles and zeros. *Int. J. Control*, **53**, 771–798. [FDM], [RPA]

Lewin, D. R. (1991). Robust performance specifications for uncertain stable SISO systems. *Int. J. Control*, **53**, 1263–1281. [HINF], [RPM], [TUTORIAL]

Lewkowitz, I. and R. Sivan (1988). Maximal stability robustness for state equations. *IEEE Trans. Aut. Control*, **AC-33**, 297–300. [SSM]

Lin, H., C. V. Hollot and A. C. Bartlett (1987). Stability of families of polynomials: geometric considerations in coefficient space. *Int. J. Control*, **45**, 649–660. [KHAR]

Lin, S. H., I. K. Fong, Y. T. Juang, T. S. Kuo and C. F. Hsu (1989). Stability of perturbed polynomials based on the argument principle and Nyquist criterion. *Int. J. Control*, **50**, 55–63. [FDM], [KHAR]

Lin, S. H., Y. T. Juang, I. K. Fong, C. F. Hsu and T. S. Kuo (1988). Dynamic interval system analysis and design. *Int. J. Control*, **48**, 1807–1818. [SSM]

Liou, C. T. and C. T. Yang (1987). Guaranteed cost control of tracking problems with large plant uncertainty. *Int. J. Control*, **45**, 2161–2171. [GCC], [RST]

Lu, W. S. (1988). 2-D stability test via 1-D stability robustness analysis. *Int. J. Control*, **48**, 1735–1741. [SSM]

Ma, C. C. H. (1990). Stability robustness of repetitive control systems with zero phase compensation. *ASME J. of Dynamic Systems, Measurement and Control*, **112**, 320–324. [RSP]

Madiwale, A. N., W. M. Haddad and D. S. Bernstein (1989). Robust H^∞ control design for systems with structured parameter uncertainty. *Systems and Control Letters*, **12**, 393–407. [HINF], [RPM], [SSM]

Magana, M. E. and S. H. Zak (1988). robust state feedback stabilization of discrete-time uncertain dynamical systems. *IEEE Trans. Aut. Control*, **AC-33**, 887–891. [DTS], [RSM], [SSM]

Magana, M. E. and S. H. Zak (1988). Robust output feedback stabilization of discrete-time uncertain dynamical systems. *IEEE Trans. Aut. Control*, **AC-33**, 1082–1085. [DTS], [RSM], [SSM]

Makila, P. M. and H. T. Toivonen (1987). Matrix criterion robust linear quadratic control problem. *Int. J. Control*, **46**, 913–931. [LQG]

Manousiouthakis, V. and Y. Arkun (1987). Hybrid approach for the design of robust control systems. *Int. J. Control*, **45**, 2203–2220. [RMM]

Martin, J. M. (1991). State space measures of robustness of pole locations for structured and unstructured perturbations. *Systems and Control Letters*, **16**, 423–433. [RPA], [SSM]

Martin, J. M. (1987). State-space measures for stability robustness. *IEEE Trans. Aut. Control*, **AC-32**, 509–512. [SSM]

Miller, D. E. and E. J. Davison (1989). The self-tuning robust servo-mechanism problem. *IEEE Trans. Aut. Control*, **AC-34**, 511–523. [RST]

Minnichelli, R. J., J. J. Anagnost and C. A. Desoer (1989). An elementary proof of Kharitonov's stability theorem with extensions. *IEEE Trans Aut. Control*, **AC-34**, 995–998. [KHAR]

Moore, J. B. and L. Xia (1987). Loop recovery and robust state estimate feedback designs. *IEEE Trans. Aut. Control*, **AC-32**, 512–517. [LQG]

Mori, T. (1989). Estimates for a measure of stability robustness via a Lyapunov matrix equation. *Int. J. Control*, **50**, 435–538. [SSM]

Mori, T. and N. Kukuma (1989). Convergence condition for convex combinations of two polynomials. *Int. J. Control*, **50**, 687–691. [KHAR]

Mori, T. and H. Kokame (1987). Convergence property of interval matrices and interval polynomials. *Int. J. Control*, **45**, 481–484. [KHAR], [SSM]

Mori, T. and H. Kokame (1990). Aperiodicity conditions for

polynomials with uncertain coefficient parameters. *Int. J. Control*, **51**, 1147–1150. [KHAR]

Mossaheb, S., R. G. Cameron and F. Li (1987). Passivity and design of sampled regulators for uncertain systems. *Int. J. Control*, **45**, 1941–1952. [DTS], [SPR]

Mustafa, D. (1989). Relations between maximum entropy H^∞ control and combined H^∞/LQG control. *Systems and Control Letters*, **12**, 193–203. [HHM], [LQG]

Mustafa, D. (1991). Combined H^∞/LQG control via the optimal projection equations: on minimizing the LQG cost bound. *Int. J. of Robust and Nonlinear Control*, **1**, 99–110. [HINF], [LQG], [RMO]

Nichols, N. K. (1987). Robustness in partial pole placement. *IEEE Trans. Aut. Control*, **AC-32**, 728–732. [RPA]

Nordstrom, K. (1987). Trade-off between noise sensitivity and robustness for LQG regulators. *Int. J. Control*, **46**, 1689–1714. [LQG], [RMO]

Nwokah, O. D. I. (1987). The robust decentralised stabilisation of complex feedback systems. *IEEE Proceedings, Part D*, **134**, 43–47. [LSS]

Nwokah, O. D. I. (1988). Quantitative design of robust multivariable control systems. *IEE Proceedings, Part D*, **135**, 57–66. [MIMO], [QFT]

Nwokah, O. D. I. and D. F. Thompson (1988). Algebraic and topological aspects of quantitative feedback theory. *Int. J. Control*, **50**, 1057–1069. [QFT]

Obinata, G. and J. B. Moore (1988). Characterization of controllers in simultaneous stabilization. *Systems and Control Letters*, **10**, 333–340. [RSP]

Okada, T., M. Kihara and M. Ikeda (1990). Robust control system design synthesis with observers. *J. of Guidance Control and Dynamics*, **13**, 337–342. [MIMO], [OBD]

Okada, T., M. Kihara and Y. Nishio (1988). Reduced conservative singular value analysis for robustness. *Int. J. Control*, **48**, 1455–1473. [MIMO]

Olbrot, A. W. (1987). Robust stabilization of uncertain systems by periodic feedback. *Int. J. Control*, **45**, 747–758. [NLC], [RSP]

Olbrot, A. W. and J. Cieslik (1988). A qualitative bound in robustness of stabilization by state feedback. *IEEE Trans. Aut. Control*, **AC-33**, 1165–1166. [RSP], [SSM]

Paice, A. D. B. and J. B. Moore (1990). Robust stabilization of nonlinear plants via left coprime factorizations. *Systems and Control Letters*, **15**, 125–135. [NTV]

Panier, E. R., M. K. H. Fan and A. L. Tits (1989). On the robust stability of polynomials with no cross-coupling between the perturbations in the coefficients of even and odd powers. *Systems and Control Letters*, **12**, 291–299. [KHAR]

Partington, J. R. and K. Glover (1990). Robust stabilization of delay systems by approximation of coprime factors. *Systems and Control Letters*, **14**, 325–331. [DPS], [RSP]

Pascoal, A. M., P. P. Khargonekar and T. T. Georgiou (1988). Pointwise stabilizability of families of linear time-invariant plants. *IEEE Trans. Aut. Control*, **AC-33**, 1161–1165. [RSP]

Petersen, I. R. (1987). Notions of stabilizability and controllability for a class of uncertain linear systems. *Int. J. Control*, **46**, 409–422. [RCO]

Petersen, I. R. (1987). A stabilization algorithm for a class of uncertain linear systems. *Systems and Control Letters*, **8**, 351–357. [RSP], [SSM]

Petersen, I. R. (1988). Stabilization of an uncertain linear system in which uncertain parameters enter into the input matrix. *SIAM J. on Control and Optimization*, **26**, 1257–1264. [NLC], [RSP], [SSM]

Petersen, I. R. (1988). Some new results on algebraic Riccati equations arising in linear quadratic differential games and the stabilization of uncertain linear systems. *Systems and Control Letters*, **10**, 341–348. [RSP], [SSM]

Petersen, I. R. (1989). A class of stability regions for which a Kharitonov-like theorem holds. *IEEE Trans. Aut. Control*, **AC-34**, 1111–1115. [KHAR]

Petersen, I. R. (1990). A new extension to Kharitonov's theorem. *IEEE Trans. Aut. Control*, **AC-35**, 825–828. [KHAR]

Petersen, I. R. and B. R. Barmish (1987). Control effort considerations in the stabilization of uncertain dynamical systems. *Systems and Control Letters*, **9**, 417–422. [NTV], [RSP]

Petkovski, D. (1987). Decentralised design of robust controllers. *IEE Proceedings, Part D*, **134**, 317–326. [LSS]

Petkovski, D. B. (1988). Stability analysis of interval matrices: improved bounds. *Int. J. Control*, **48**, 2265–2273. [SSM], [KHAR]

Pohjolainen, S. (1987). On the optimal tuning of a robust controller for parabolic distributed parameter systems. *Automatica*, **23**, 719–728. [DPS], [SSM]

Poolla, K. and T. Ting (1987). Nonlinear time-varying controllers for robust stabilization. *IEEE Trans. Aut Control*, **AC-32**, 195–200. [NLC], [RSP]

Porter, B. and A. Khaki-Sedigh (1989). Robustness properties of tunable digital set-point tracking PID controllers for linear multivariable plants. *Int. J. Control*, **49**, 777–789. [MIMO], [RST]

Pritchard, A. J. and S. Townley (1988). Robust compensator design via structured stability radii. *Systems and Control Letters*, **11**, 33–37. [SSM]

Pujara, L. R. (1990). On the stability of uncertain polynomials with dependent coefficients. *IEEE Trans. Aut Control*, **AC-35**, 756–759. [KHAR]

Qui, L. and E. J. Davison (1989). A simple procedure for the exact stability robustness computation of polynomials with affine coefficient perturbations. *Systems and Control Letters*, **13**, 413–420. [KHAR]

Qui, L. and E. J. Davison (1991). The stability robustness determination of state space models with real unstructured perturbations. *Mathematics of Control, Signal and Systems*, **4**, 247–267. [SSM]

Rachid, A. (1989). Robustness of discrete systems under structured uncertainties. *Int. J. Control*, **50**, 1563–1566. [DTS]

Rantzer, A. (1990). Hurwitz testing sets for parallel polytopes of polynomials. *Systems and Control Letters*, **15**, 99–104. [KHAR]

Rotea, M. A. and P. P. Khargonekar (1988). Stabilizability of linear time-varying and uncertain linear systems. *IEEE Trans Aut Control*, **AC-33**, 884–887. [NTV], [RSP]

Rotea, M. A. and P. P. Khargonekar (1989). Stabilization of uncertain systems with norm bounded uncertainty—A control Lyapunov function approach. *SIAM J. on Control and Optimization*, **27**, 1462–1476. [SSM], [RSP]

Rotea, M. A. and P. P. Khargonekar (1991). H^2-optimal control with an H^∞-constraint: the state feedback case. *Automatica*, **27**, 307–316. [HHM]

Rotstein, H., A. Desages and J. A. Romagnoli (1989). Robustness consideration in reduced-order controllers. *IEEE Trans. Aut. Control*, **AC-34**, 457–459. [ROD]

Rotstein, H., A. Desages and J. Romagnoli (1989). Calculation of highly structured stability margins. *Int. J. Control*, **49**, 1079–1092. [SSV]

Rotstein, H., R. Sanchez Pena, J. Bandoni, A. Desages and J. Romagnoli (1991). Robust characteristic polynomial assignment. *Automatica*, **27**, 711–716. [RPA]

Saberi, A. and P. Sannuti (1988). Global stabilization with almost disturbance decoupling of a class of uncertain non-linear systems. *Int. J. Control*, **47**, 717–727. [NTV]

Saberi, A. and P. Sannuti (1990). Observer design for loop-transfer recovery and for uncertain dynamical systems. *IEEE Trans. Aut. Control*, **AC-35**, 878–897. [LTR], [OBD]

Saberi, A. and P. Sannuti (1990). Observer-based control of uncertain systems with nonlinear uncertainties. *Int. J. Control*, **52**, 1107–1130. [NTV], [OBD]

Safonov, M. G., D. J. N. Limebeer and R. Y. Chiang (1989). Simplifying the H^∞ theory via loop-shifting, matrix-pencil and descriptor concepts. *Int. J. Control*, **50**, 2467–2488. [HINF]

Sanchez Pena, R. S. and A. Sideris (1990). Robustness with real parametric and structured complex uncertainty. *Int. J. Control*, **52**, 753–765. [SSV]

Santos-Mendes, R. and J. Aguilar-Martin (1989). Robust pole placement design. *Int. J. Control*, **50**, 113–128. [RPA]

Saydy, L., A. L. Tits and E. H. Abed (1990). Guardian maps and the generalized stability of parametrized families of matrices and polynomials. *Mathematics of Control, Signals and Systems*, **3**, 345–371. [KHAR], [SSM]

Saydy, L., A. L. Tits and E. H. Abed (1991). Maximal range for generalized stability—applications to physically motivated examples. *Int. J. Control*, **53**, 837–845. [KHAR], [SSM]

Schiavoni, N. (1990). Further results on the design of multivariable regulators under structure and robustness constraints. *Int. J. Control*, **51**, 669–685. [MIMO]

Schmitendorf, W. E. (1987). Methods for obtaining robust tracking control laws. *Automatica*, **23**, 675–677. [RST]

Schmitendorf, W. E. (1988). Design of observer-based robust stabilizing controllers. *Automatica*, **24**, 693–696. [OBD], [RSP]

Schmitendorf, W. E. (1988). Designing stabilizing controllers for uncertain systems using the Riccati equation aproach. *IEEE Trans. Aut. Control*, **AC-33**, 376–379. [SSM]

Schmitendorf, W. E. and C. V. Hollot (1989). Simultaneous stabilization via linear state feedback control. *IEEE Trans. Aut. Control*, **AC-34**, 1001–1005. [RSP], [SSM]

Sezer, M. E. and D. D. Siljak (1988). Robust stability of discrete systems. *Int. J. Control*, **48**, 2055–2063. [DTS]

Shafai, B., K. Perev, D. Crowley and Y. Chehab (1991). A necessary and sufficient condition for the stability of nonnegative interval discrete systems. *IEEE Trans. Aut. Control*, **AC-36**, 742–746. [DTS], [KHAR], [SSM]

Shamma, J. S. and M. A. Dahleh (1991). Time-varying versus time-invariant compensation for rejection of persistent bounded disturbances and robust stabilization. *IEEE Trans. Aut. Control*, **AC-36**, 838–847. [HINF], [LONE], [NLC], [NTV]

Sideris, A. (1991). An efficient algorithm for checking the robust stability of a polytope of polynomials. *Mathematics of Control, Signals and Systems*, **4**, 315–337. [KHAR]

Sideris, A. and B. R. Barmish (1989). An edge theorem for polytopes of polynomials which can drop in degree. *Systems and Control Letters*, **13**, 232–238. [KHAR]

Sideris, A. and R. S. Sanchez-Pena (1990). Robustness margin calculation with dynamic and real parametric uncertainty. *IEEE Trans. Aut. Control*, **AC-35**, 970–974. [SSV]

Sideris, A. and R. S. Sanchez-Pena (1989). Fast computation of the multivariable stability margin for real interrelated uncertain parameters. *IEEE Trans Aut. Control*, **AC-34**, 1272–1276. [KHAR]

Siljak, D. D. (1989). Parameter space methods for robust control design: a guided tour. *IEEE Trans. Aut. Control*, **AC-34**, 674–688. [KHAR], [SURVEY]

Shi, Z. C. and W. B. Gao (1987). Stability of interval parameter matrices. *Int. J. Control*, **45**, 1093–1101. [SSM]

Sobel, K. M., S. S. Banda and H. H. Yeh (1989). Robust control for linear systems with structured state space uncertainty. *Int. J. Control*, **50**, 1991–2004. [SSM]

Sogaard-Andersen, P. (1987). Comments on the loop transfer recovery. *Int. J. Control*, **45**, 369–374. [LTR]

Soh, C. B. (1990). Necessary and sufficient conditions for stability of symmetric interval matrices *Int. J. Control*, **51**, 243–248. [KHAR]

Soh, C. B. and C. K. Khan (1989). Destabilizing perturbations for linear systems. *Int. J. Control*, **50**, 2101–2104. [SSM]

Soh, Y. C. (1989). Robust pole-placement for uncertain interval systems. *IEE Proceedings, Part D*, **136**, 301–306. [RPA]

Soh, Y. C. (1989). Zero locations of an entire family of polytope polynomials. *Int. J. Control*, **49**, 1851–1859. [KHAR]

Soh, Y. C. (1989). Strict Hurwitz property of polynomials under coefficient perturbation. *IEEE Trans. Aut. Control*, **AC-34**, 629–632. [KHAR]

Soh, Y. C. and R. J. Evans (1988). Stability analysis of interval matrices-continuous and discrete systems. *Int. J. Control*, **47**, 25–32. [DTS], [KHAR], [SSM]

Soh, Y. C., R. J. Evans, I. R. Petersen and R. E. Betz (1987). Robust pole assignment. *Automatica*, **23**, 601–610. [RPA]

Soh, Y. C. and Y. K. Foo (1989). Generalized edge theorem. *Systems and Control Letters*, **12**, 219–224. [KHAR]

Soh, Y. C. and Y. K. Foo (1990). Generalization of strong Kharitonov's theorems to the left sector. *IEEE Trans. Aut. Control*, **AC-35**, 1378–1382. [KHAR]

Soh, Y. C., Y. K. Foo and C. B. Soh (1989). Perturbed polynomials with zeros in the left sector. *Systems and Control Letters*, **12**, 415–419. [KHAR]

Stein, G. and J. C. Doyle (1991). Beyond singular values and loop shapes. *J. of Guidance, Control and Dynamics*, **14**, 5–16. [LTR], [SSV], [TUTORIAL]

Sugie, T. and M. Vidyasagar (1989). Further results on the robust tracking problem in two-degree-of-freedom control systems. *Systems and Control Letters*, **13**, 101–108. [RST]

Sung, H. K. and S. Hara (1988). Properties of sensitivity and complementary sensitivity functions in SISO digital control systems. *Int. J. Control*, **48**, 2429–2439. [DTS], [HINF], [MSN]

Sung, H. K. and S. Hara (1989). Properties of complementary sensitivity function in SISO digital control systems. *Int. J. Control*, **50**, 1283–1295. [DTS]

Tadmor, G. (1990). Worst-case design in the time-domain: the maximum principle and the standard H^∞ problem. *Mathematics of Control, Signal and Systems*, **3**, 301–324. [HINF], [SSM]

Tadmor, G. (1991). Uncertain feedback loops and robustness in general linear systems. *Automatica*, **27**, 1039–1042. [DPS], [HINF]

Tempo, R. (1990). A dual result to Kharitonov's theorem. *IEEE Trans. Aut. Control*, **AC-35**, 195–198. [KHAR]

Tesi, A. and A. Vicino (1989). A new fast algorithm for robust stability analysis of control systems with linearly dependent parametric uncertainties. *Systems and Control Letters*, **13**, 321–329. [KRF]

Tesi, A. and A. Vicino (1990). Robustness analysis for linear dynamical systems with correlated parametric uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 186–191. [KHAR], [KRF]

Tesi, A. and A. Vicino (1990). Robust stability of state-space models with structured uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 191–195. [KHAR], [SSM]

Tesi, A. and A. Vicino (1991). Robust absolute stability of Lur'e control systems in parameter space. *Automatica*, **27**, 147–151 [NTV], [KHAR]

Ting, T. and K. Poolla (1988). Upper bounds and approximate solutions for multidisk problems. *IEEE Trans. Aut. Control*, **AC-33**, 783–786. [HINF], [RMO]

Townley, S. (1990). Stability radius optimization: a geometric approach. *Systems and Control Letters*, **14**, 199–207. [SSM]

Townley, S. and E. P. Ryan (1991). On the supremally regulated complex stability radius as a real robustness measure. *Systems and Control Letters*, **16**, 249–259. [SSM]

Tsai, M. C and Y. P. Shih (1989). H^∞ optimal robust control of feedback systems. *ASME J. of Dynamic Systems, Measurement and Control*, **111**, 336–339. [HINF]

Tsai, T. P. and T. S. Wang (1987). Optimal design of non-minimum-phase control systems with large plant uncertainty. *Int. J. Control*, **45**, 2147–2159. [FDM]

Tsay, S. C. (1990). Robust control for linear uncertain systems via linear quadratic state feedback. *Systems and Control Letters*, **15**, 199–205. [LQG]

Tsay, S. C., I. K. Fong and T. S. Kuo (1991). Robust linear quadratic optimal control for systems with linear uncertainties. *Int. J. Control*, **53**, 81–96. [LQG], [RPM]

Tsypkin, Y. Z. and B. T. Polyak (1991). Frequency domain criteria for l^p Robust stability of continuous linear systems. *IEEE Trans. Aut Control*, **AC-36**, 1464–1469. [FDM], [KHAR]

Tsui, C. C. (1987). On preserving the robustness of an optimal control system with observers. *IEEE Trans. Aut. Control*, **AC-32**, 823–826. [OBD]

Van de Vegte, J. (1991). Sequential design of multivariable

control systems. *IEE Proceedings, Part D*, **138**, 103–110. [MIMO]
Vardulakis, A. I. G. (1988). Necessary and sufficient condition for stability robustness under known additive perturbation. *Int. J. Control*, **47**, 1259–1266. [MIMO]
Vaz, A. F. and E. J. Davison (1988). The decentralized control expanding system problem. *Systems and Control Letters*, **10**, 301–308. [LSS]
Vaz, A. F. and E. J. Davison (1989) The structured robust decentralized servomechanism problem for interconnected systems. *Automatica*, **25**, 267–272. [LSS], [RST]
Verma, M. S. (1989). Robust stabilization of linear-time-invariant systems. *IEEE Trans. Aut. Control*, **AC-34**, 870–875. [HINF]
Vicino, A. (1988). Some results on robust stability of discrete-time systems. *IEEE Trans. Aut. Control*, **AC-33**, 844–847. [DTS], [KHAR]
Vicino, A. (1989). Robustness of pole location in perturbed systems. *Automatica*, **25**, 109–113. [KRF]
Vicino, A. (1989). Maximal polytopic stability domains in parameter space for uncertain systems. *Int. J. Control*, **49**, 351–361. [KRF]
Vicino, A., A. Tesi and M. Milanese (1990). Computation of nonconservative stability perturbation bounds for systems with nonlinearly correlated uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 835–841. [KRF]
Walker, D. J. (1990). Robust stabilizability of discrete-time systems with normalized stable factor perturbation. *Int. J. Control*, **52**, 441–455. [DTS]
Wang, S. D. and T. S. Kuo (1989). Performance bounds of optimal linear systems with uncertain parameters. *Int. J Control*, **50**, 1185–1194. [RPM]
Wang, S. D. and T. S. Kuo (1990). Design of robust linear state feedback laws: ellipsoidal set-theoretic approach. *Automatica*, **26**, 303–309. [SSM]
Wang, S. D., T. S. Kuo and C. F. Hsu (1987). Optimal-observer design for linear dynamical systems with uncertain parameters. *Int. J. Control*, **45**, 701–711. [OBD]
Wang, S. D., T. S. Kuo, Y. H. Lin, C. F. Hsu and Y. T Juang (1987). Robust control design for linear systems with uncertain parameters. *Int. J. Control*, **46**, 1557–1567 [QSM]
Wang, S. S., B. S. Chen and T. P. Lin (1987). Robust stability of uncertain time-delay systems. *Int J. Control*, **46**, 963–976. [DPS]
Wang, T. S., Y. C. Hwang and T. P Tasi (1990). Synthesis of feedback control systems with large plant uncertainty and time-varying gain. *Int. J. Control*, **51**, 687–704. [NTV], [QFT]
Wang, W. J. and C. F. Cheng (1989). Robustness of perturbed large-scale systems with local constant state feedback. *Int. J. Control*, **50**, 373–384. [LSS]
Wei, K. H. and B. R. Barmish (1989). Making a polynomial Hurwitz-invariant by choice of feedback gains *Int. J. Control*, **50**, 1025–1038. [KHAR]
Wei, K. H. and R. K. Yedavalli (1989). Robust stabilizability for linear systems with both parameter variation and unstructured uncertainty. *IEEE Trans. Aut. Control*, **AC-34**, 149–156. [MSU], [RSP]
Wei, K. (1990). Stabilization of a linear plant via a stable compensator having no real unstable zeros. *Systems and Control Letters*, **15**, 259–264. [RSP]
Wu, Q. H. and M. Mansour (1989). Robust output regulation for a class of linear multivariable systems. *Systems and Control Letters*, **13**, 227–232. [HINF], [MIMO]
Wu, Q. H. and M. Mansour (1990). H^∞-optimal solutions of robust regulator problem for linear MIMO systems. *Int. J. Control*, **52**, 1241–1262. [HINF], [MIMO]
Xin, L. X. (1987). Stability of interval matrices *Int. J. Control*, **45**, 203–210. [KHAR], [SSM]
Xin, L. X. (1987). Necessary and sufficient conditions for stability of a class of interval matrices. *Int. J. Control*, **45**, 211–214. [KHAR], [SSM]
Xie, L. and C. E. de Souza (1990). Robust H^∞ control for linear time-invariant systems with norm-bounded uncertainty in the input matrix. *Systems and Control Letters*, **14**, 389–396. [HINF]
Xie, L. and C. E. de Souza (1991). Robust H^∞ control for class of uncertain linear time-invariant systems. *IEE Proceedings, Part D*, **138**, 479–483. [HINF], [QSM]
Yan, W. and B. D. O. Anderson (1990). The simultaneous optimization problem for sensitivity and gain margin. *IEEE Trans. Aut. Control*, **AC-35**, 558–563. [RMO]
Yang, W C. and M. Tomizuka (1990). Discrete time robust control via state feedback for single input systems. *IEEE Trans. Aut. Control*, **AC-35**, 590–598. [DTS]
Yaniv, O. (1991). Robust design of MIMO feedback systems having an uncertain non-linear plant. *Int J. Control*, **53**, 1283–1294. [MIMO], [QFT]
Yaniv, O. (1991). Arbitrarily small sensitivity in MIMO uncertain feedback systems. *Automatica*, **27**, 565–568. [MIMO], [QFT]
Yaniv, O., P O Gutman and L Neumann (1990) An algorithm for the adaptation of a robust controller to reduced plant uncertainty. *Automatica*, **26**, 709–720 [QFT]
Yaniv, O. and B. Schwartz (1991). Criterion for loop stability in MIMO feedback systems having an uncertain plant. *Int. J. Control*, **53**, 527–539. [MIMO], [QFT]
Yaz, E. (1988) Deterministic and stochastic robustness measures for discrete systems. *IEEE Trans Aut. Control*, **AC-33**, 952–955 [DTS]
Yaz, E. and X. Niu (1989). Stability robustness of linear discrete-time systems in the presence of uncertainty. *Int. J. Control*, **50**, 173–182 [DTS], [SSM]
Yeh, H. H., S S. Banda and P. J. Lynch (1990) Control of unknown systems via deconvolution. *J. of Guidance, Control and Dynamics*, **13**, 416–424. [SSV]
Yeh, H. H, S. S. Banda, P. J. Lynch and T. E. McQuade (1988). Loop-transfer recovery via Hardy-space optimization. *J. of Guidance, Control and Dynamics*, **11**, 86–90. [LTR]
Zak, S H (1990). On the stabilization and the observation of non-linear uncertain dynamic systems. *IEEE Trans. Aut. Control*, **AC-35**, 604–607. [NTV]
Zeheb, E (1989) Necessary and sufficient conditions for root clustering of a polytope of polynomials in a simply connected domain. *IEEE Trans. Aut Control*, **AC-34**, 986–990. [KHAR]
Zeheb, E. (1990). Necessary and sufficient conditions for robust stability of a continuous system—the continuous dependency case illustrated via multilinear dependency. *IEEE Trans. Circ and Systems*, **CAS-37**, 47–53 [KHAR]
Zhao, Y. and H. Kimura (1988). Two-degree-of-freedom dead-beat control system with robustness. *Int. J. Control*, **48**, 303–315. [RSP]
Zhao, Y. and H. Kimura (1989) Two-degree-of-freedom dead-beat control system with robustness: multivariable case. *Int. J. Control*, **49**, 667–679 [MIMO], [RSP]
Zhou, K. and P P. Khargonekar (1988). On the stabilization of uncertain linear systems via bound invariant Lyapunov functions. *SIAM J. on Control and Optimization*, **26**, 1265–1273 [QSM], [RSP], [SSM]
Zhou, K. and P. P. Khargonekar (1988). Robust stabilization of linear systems with norm-bounded time-varying uncertainty. *Systems and Control Letters*, **10**, 17–20. [NTV], [QSM], [RSP], [SSM]
Zhu, S. Q., M. L J. Hautus and C. Praagman (1988). Sufficient conditions for robust BIBO stabilization: given by the gap metric. *Systems and Control Letters*, **11**, 53–59. [HINF], [RSP]
Zohdy, M. A., N. K. Loh and A. S. Abdul-Wahab (1987). A robust optimal model matching control. *IEEE Trans. Aut. Control*, **AC-32**, 410–414. [RMM]

Journal articles: Applications

DC-motors

Eisaka, T. Y. Zhong, S. Bai and R. Tagawa (1989). Evaluation of robust model matching for the control of a DC servo motor. *Int. J. Control*, **50**, 479–493. [RMM]

Horowitz, R., H. I. Stephens and G. Leitmann (1989). Experimental implementation of a deterministic controller for a DC motor with uncertain dynamics. *ASME J. of Dynamic Systems, Measurement and Control*, **111**, 244–252. [NTV]

Stephan, R. M., V. Hahn, J. Dastych and H. Unbehauen (1991). Adaptive and robust cascade schemes for thyristor driven DC-motor speed control. *Automatica*, **27**, 449–461. [DTS]

Electromechanical

Bada, A. T. (1987). Robust brake control for a heavy-duty truck. *IEE Proceedings, Part D*, **134**, 1–8. [RMO]

Salman, M. A. (1990). A robust servo-electronic controller for brake force distribution. *ASME J. of Dynamic Systems, Measurement and Control*, **112**, 442–447. [FDM]

Engine control

Kuraoka, Hiroaki, Naoto Ohka, Masahiro Ohba, Shigeyuki Hosoe and Feifei Zhang (1990). Application of H^∞ design to automobile fuel control. *IEEE Control Systems Magazine*, Apr. 1990, 102–106. [HINF]

Flexible structures

Balas, Gary J. and John C. Doyle (1990). Identification of flexible structures for robust control. *IEEE Control Systems Magazine*, Jun. 1990, 51–58. [MOD], [SSV]

Blelloch, P. A. and D. L. Mingrori (1990). Robust linear quadratic Gaussian control for flexible structures. *J. of Guidance, Control and Dynamics*, **13**, 66–72. [LQG]

Bontsema, J. and R. F. Curtain (1988). A note on spillover and robustness for flexible systems. *IEEE Trans. Aut. Control*, **AC-33**, 567–569. [HINF], [LQG]

Bontsema, J., R. F. Curtain and J. M. Schumacher (1988). Robust control of flexible structures: a case study. *Automatica*, **24**, 177–186. [HINF]

Byun, K. W., B. Wie, D. Geller and J. Sunkel (1991). Robust H^∞ control design for the state space solution with structured real parameters. *J. of Guidance, Control and Dynamics*, **14**, 1115–1122. [HINF]

Collins, Emmanuel G. Jr., Douglas J. Phillips and David C. Hyland (1991). Robust decentralized control laws for the ACES structure. *IEEE Control Systems Magazine*, Apr. 1991, 62–70. [ROD], [SSM]

Kimura, H., T. Oike, A. Miura, K. Akai and T. Kida (1991). Robust stability-Degree assignment and its applications to the control of flexible structures. *Int. J. of Robust and Nonlinear Control*, **1**, 153–170. [KHAR]

Lim, Kyong B. and John L. Junkins (1989). Robustness optimization of structural and controller parameters. *J. of Guidance, Control, and Dynamics*, **12**, 89–96. [SSM]

Opdenacker, P. C., E. A. Jonckheere, M. G. Safonov, J. C. Juang and M. S. Lukich (1990). Reduced-order compensator design for a flexible structure. *J. of Guidance, Control and Dynamics*, **13**, 46–56. [HINF], [ROD]

Rew, D. W., J. L. Junkins and J. N. Juang (1989). Robust eigenstructure assignment by a projection method: applications using multiple optimization criteria. *J. of Guidance, Control and Dynamics*, **12**, 396–403. [RMO], [RPA]

Safonov, M. G., R. Y. Chiang and H. Flashner (1991). H^∞ robust control synthesis for a large space structure. *J. of Guidance, Control and Dynamics*, **14**, 513–520. [HINF]

Tahk, Minjea and Jason L. Speyer (1989). Parameter robust linear-quadratic-Gaussian design synthesis with flexible structure control applications. *J. of Guidance, Control and Dynamics*, **12**, 460–468. [LQG], [LTR]

Flight control

Apkarian, Pierre R. (1989). Structured stability robustness improvement by eigenspace techniques: a hybrid methodology. *J. of Guidance, Control and Dynamics*, **12**, 162–168. [SSV]

Arrow, A. and D. E. Williams (1989). Comparison of classical and modern autopilot design and analysis techniques. *J. of Guidance, Control, and Dynamics*, **12**, 220–228. [LTR]

Enns, Dale F. (1991). Rocket stabilization as a structured singular value synthesis design example. *IEEE Control Systems Magazine*, Jun. 1991, 67–73. [RSP], [SSV]

Hopp, T. H. and W. E. Schmitendorf (1990). Design of a linear controller for robust tracking and model following. *ASME J. of Dynamic Systems, Measurement and Control*, **112**, 552–558. [QSM], [RSP]

Kaminer, Isaac, Pramod P. Khargonekar and Greg Robel (1990). Design of localized capture and track modes for a lateral autopilot using H^∞ synthesis. *IEEE Control Systems Magazine*, Jun. 1990, 13–21. [HINF]

McRuer, Duane T., Thomas T. Myers and Peter M. Thompson (1989). Later singular-value-based flight control system design techniques. *J. of Guidance, Control and Dynamics*, **12**, 913–919. [MIMO], [MSU], [SSV]

Mudge, S. K. and R. J. Patton (1988). Analysis of the technique of robust eigenstructure assignment with application to aircraft control. *IEE Proceedings, Part D*, **135**, 275–281. [RPA]

Mukhopadhyay, Vivek (1989). Digital robust control law synthesis using constrained optimization. *J. of Guidance, Control and Dynamics*, **12**, 175–181. [DTS], [QSM]

Okada, Tsuyoshi, Masahiko Kihara and Masakazu Ikeda (1990). Robust control system design synthesis with observers. *J. of Guidance, Control and Dynamics*, **13**, 337–342. [MIMO], [OBD]

Petkovski, Djordjija B. (1989). Improved time-domain stability robustness measures for linear regulators. *J. of Guidance, Control and Dynamics*, **12**, 595–597. [SSM]

Safonov, M. G. and R. Y Chiang (1988). CACSD using the state-space L^∞ theory—A design example. *IEEE Trans. Aut. Control*, **AC-33**, 477–479. [HINF]

Wise, K. A. (1991). Singular value robustness tests for missile autopilot uncertainties. *J. of Guidance, Control and Dynamics*, **14**, 597–606. [SSV]

White, J. E. and J. R. Etter (1990). Linear-quadratic regulator pointing system for high-altitude balloon payload. *J. of Guidance, Control and Dynamics*, **13**, 615–623. [LQG], [MIMO]

Yue, A. and I. Postlethwaite (1990) Improvement of helicopter handling qualities using H^∞-optimisation. *IEE Proceedings, Part D*, **137**, 115–129. [HINF].

Mechanical systems

Byun, K. W., B. Wie and J. Sunkel (1991). Robust non-minimum phase compensation for a class of uncertain dynamical systems. *J. of Guidance, Control and Dynamics*, **14**, 1191–1199. [HINF], [QFT]

Guttalu, R. S. and J. M. Skowronski (1989). Algorithms for degree of controllability of uncertain mechanical systems. *ASME J. of Dynamic Systems, Measurement and Control*, **111**, 600–604. [NTV], [SSM]

Keel, L. H., K. B. Lim and J. N. Juang (1991). Robust eigenvalue assignment with maximum tolerance to system uncertainties. *J. of Guidance, Control and Dynamics*, **14**, 615–620. [RPA], [SSM]

Wedell, E., C. H. Chuang and B. Wie (1991). Stability robustness margin computation for structured real-parameters. *J. of Guidance, Control and Dynamics*, **14**, 607–614. [FDM], [KHAR]

Process control

Lee, Jay H. and Manfred Morari (1991). Robust measurement selection. *Automatica*, **27**, 519–527. [SSV]

Skogestad, S., M. Morari and J. C. Doyle (1988). Robust control of ill-conditioned plants: high-purity distillation. *IEEE Trans. Aut. Control*, **AC-33**, 1092–1105. [HINF], [MIMO], [SSV]

Robotics

Abdallah, C., C. Dawson, P. Dorato and M. Jamishidi (1991). Survey of robust control for rigid bodies. *IEEE*

Control Systems Magazine, Feb 1991, 24–30 [NTV], [SSM], [SURVEY]

Becker, Norbert and Wolfgang M Grimm (1988) On L_2 and L_∞ stability approaches for the robust control of robot manipulators *IEEE Trans Aut Control*, **AC-33**, 118–122 [HINF]

Canudas de Wit, C and N Fixot (1991) Robot control via robust estimated state feedback *IEEE Trans Aut Control*, **AC-36**, 1497–1501 [SSM]

Corless, M (1989) Tracking controllers for uncertain systems applications to a Manutec R3 robot *ASME J of Dynamic Systems, Measurement and Control* **111**, 609–618 [NTV]

Kuo, C Y and Shay-Ping T Wang (1989) Nonlinear robust industrial robot control *ASME J of Dynamic Systems Measurement and Control*, **111**, 24–30 [NTV]

Kuo, C Y and Shay-Ping T Wang (1990) Nonlinear robust hybrid control of robotic manipulators *ASME J of Dynamic Systems, Measurement and Control* **111**, 48–54 [NTV]

Mills, J K and A A Goldenberg (1989) Robust control of robotic manipulators in the presence of dynamic parameter uncertainty *ASME J of Dynamic Systems, Measurement and Control*, **111**, 444–451 [RSP], [SSM]

Qu, Z , J F Dorsey, X Zhang and D M Dawson (1991) Robust control of robots by the computed torque law *Systems and Control Letters*, **16**, 25–32 [NTV], [QSM], [RSP]

Shin, K G and N D McKay (1987) Robust trajectory planning for robotic manipulators under payload uncertainties *IEEE Trans Aut Control*, **AC-32**, 1044–1054 [NTV]

Shoureshi, Rahmat, Michael E Momot and M D Roesler (1990) Robust control for manipulators with uncertain dynamics *Automatica*, **26**, 353–359 [NTV] [SSM]

Spong, M W , J S Thorp and J M Kleinwaks (1987) Robust microprocessor control of robot manipulators *Automatica*, **23**, 373–379 [DTS], [NTV]

Thermal systems

Gueslaga, A R and H W Kropholler (1990) Improved temperature and humidity control using H^∞ synthesis *IEE Proceedings, Part D*, **137**, 374 280 [HINF]

Automatica, Vol. 29, No. 1, pp. 215–217, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

On Robust Stability of Polynomials with Polynomial Parameter Dependency: Two/three Parameter Cases*

D. KAESBAUER†

Key Words—Robust control; robustness; stability, polynomials; nonlinear equations.

Abstract—We consider real polynomials whose coefficients depend polynomially on the elements of an uncertain parameter vector. The size of perturbation is characterized by the weighted norm of the parameter vector. The smallest destabilizing perturbation defines the stability radius of the set of uncertain polynomials.

It is shown that determining this radius is equivalent to solving a finite set of systems of algebraic equations and picking out the real solution with the smallest norm. The number of systems of equations depends crucially on the dimension of the parameter vector, whereas the complexity of systems of equations increases mainly with the kind of polynomial dependency and the degree of the polynomial. This method also yields the smallest destabilizing parameter combination and the corresponding critical frequency For two or three parameters this transformed problem can be solved using symbolic and numeric computations

1. *Introduction*

CONSIDER A CONTROL system with uncertain real parameters. Assume that the controller was designed for some nominal value of the parameters. How far can the parameters be changed until the system becomes unstable? A crucial role in solving this problem is the type of dependency of the coefficients of the characteristic polynomial. For linear (affine) dependency there exist several methods to determine the stability radius. There are both graphical and analytical procedures by Barmish (1989), Chapellat and Bhattacharyya (1988), Kaesbauer and Ackermann (1990) and Polyak and Tsypkin (1990). A much more difficult problem arises when we assume multilinear or polynomial dependency. Vicino *et al.* (1990) transform the problem into an optimization problem and present a numerical algorithm to find the solution. Murdock and Schmittendorf (1991) solve the same problem using a genetic algorithm. We will show that the problem can also be solved in an analytical–numerical way that is we have to solve systems of algebraic equations. We first demonstrate how to derive these systems of equations and then how to find the solutions.

2. *Problem formulation and basic results*

Given a family of polynomials

$$P(s,\mathbf{q}) = a_n(\mathbf{q})s^n + \cdots + a_1(\mathbf{q})s + a_0(\mathbf{q}), \tag{1}$$

with $\mathbf{q} \in \mathcal{R}^l$ and $a_i(\mathbf{q}) \in \mathcal{R}[q_1, q_2, \ldots, q_l]$ $(i = 0, 1, \ldots, n)$. The nominal polynomial $P(s, \mathbf{0})$ is assumed stable. Find the maximal ρ s.t. $P(s, \mathbf{q})$ is stable for all $\|\mathbf{q}\|_p < \rho$. ρ is called the stability radius.

Concerning the choice of the norm, i.e. p, there are three important possibilities. For $p = \infty$ the set of admissible $\mathbf{q}$ describes an l-dimensional cube. Dual to this norm is $p = 1$, which corresponds to a diamond. $p = 2$ yields an l-dimensional sphere in q-space. From the practical point of view the case $p = \infty$ is the most important one, because there the uncertain parameters vary independently. We will handle the case $p = \infty$ in detail. For the other cases it is not difficult to derive the corresponding results.

It is known that the stable set of $\mathbf{q}$ is bounded by three hypersurfaces, namely $a_0(\mathbf{q}) = 0$, $a_n(\mathbf{q}) = 0$ and $H_{n-1}(\mathbf{q}) = 0$. The last equation is the last but one Hurwitz determinant this is the critical one. It results from the elimination of ω from the two equations $\operatorname{Re} P(j\omega) = 0$ and $\operatorname{Im} P(j\omega) = 0$. Here linear or nonlinear dependency make a big difference. For fixed ω $\operatorname{Re} P = 0$, $\operatorname{Im} P = 0$ represent in case of linear dependency a linear manifold; this means for example that for $l = 3 H_2(\mathbf{q}) = 0$ is generated by the continuous movement of a straight line. In the case of nonlinear dependency we have a set of curves in $\mathcal{R}^3$, which generate $H_2 = 0$. Evaluating $H_{n-1}(\mathbf{q})$ must in general already be done by symbolic computations.

All three equations of the hypersurfaces will be treated in the same way The only difference is that the third one will be the complicated one with respect to the number of terms and degree of the terms. In the sequel we use $F(\mathbf{q}) = 0$ for each of these equations.

3. *Two or three parameters*

We consider first the case of two parameters, where we can demonstrate the basic idea. The polynomial family with $\|\mathbf{q}\|_\infty \le \epsilon$ (ϵ sufficient small) is stable and can be described by a square of side length 2ϵ We enlarge this square continuously until there is a point of intersection with the curve $F(q_1, q_2) = 0$. This point may lie on a vertex or on an edge of the square. The first situation is characterized by the fact that $q_1 = q_2$ or $q_1 = -q_2$, which results in the two polynomials

$$F(q_1, q_1) = 0, \quad F(q_1, -q_1) = 0. \tag{2}$$

The second case is an intersection point on an edge. This means that $F(q_1, q_2) = 0$ has a horizontal or a vertical tangent. This necessary condition leads to the two systems of equations in two unknowns

$$\begin{aligned} F(q_1, q_2) = 0, &\quad \frac{\partial F(q_1, q_2)}{\partial q_1} = 0, \\ F(q_1, q_2) = 0, &\quad \frac{\partial F(q_1, q_2)}{\partial q_2} = 0. \end{aligned} \tag{3}$$

It may be possible that in the intersection point the curve cannot be differentiated, i.e. the curve has for example a cusp and both partial derivatives vanish, but this is already covered by (3). Solving the two polynomials and the two systems of equations give us a set of points (q_1, q_2). The

* Received 3 September 1991; received in final form 17 March 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor P. Dorato.

† DLR, Oberpfaffenhofen, Germany.

point with the smallest norm determines the stability radius. For the first case we have $\|(\pm q, \pm q)\|_\infty = |q|$ and for the second case $\|(q_1, q_2)\|)_\infty = \max(|q_1|, |q_2|)$. The critical parameter combinations (q_1, q_2) also fixes the critical frequency. The polynomial $P(s, q_1, q_2) = 0$ has a root at $s = 0$ or at $s = \infty$ or a root pair at $s = \pm j\omega$. All other roots are not in the right half plane.

Consider now the case of three parameters. Instead of a curve we now have a surface $F(q_1, q_2, q_3) = 0$ which bounds the stable polynomials and a cube is enlarged now. The intersection points lie either on a vertex, on an edge or on an surface of the cube. Corresponding to the eight vertices of the cube we have the four polynomials

$$
\begin{aligned}
F(q, +q, +q) &= 0, \\
F(q, +q, -q) &= 0, \\
F(q, -q, +q) &= 0, \\
F(q, -q, -q) &= 0.
\end{aligned}
$$

In the case of 12 edges, two intersection points must coincide, i.e. the partial derivatives must vanish and we have to solve the six systems of equations

$$
\begin{aligned}
&F(q_1, +q, +q) = 0, && \left.\frac{\partial F}{\partial q_1}\right|_{q_2=+q, q_3=+q} = 0, \\
&F(q_1, +q, -q) = 0, && \left.\frac{\partial F}{\partial q_1}\right|_{q_2=+q, q_3=-q} = 0, \\
&F(+q, q_2, +q) = 0, && \left.\frac{\partial F}{\partial q_2}\right|_{q_1=+q, q_3=+q} \\
&F(-q, q_2, +q) = 0, && \left.\frac{\partial F}{\partial q_2}\right|_{q_1=-q, q_3=+q} \\
&F(+q, +q, q_3) = 0, && \left.\frac{\partial F}{\partial q_3}\right|_{q_1=+q, q_2=+q} = 0, \\
&F(+q, -q, q_3) = 0, && \left.\frac{\partial F}{\partial q_3}\right|_{q_1=+q, q_2=-q} = 0,
\end{aligned}
$$

in two unknowns.

If the intersection is on one of the six surfaces then the normal vector of the surface $F(q_1, q_2, q_3) = 0$ is parallel to one of the coordinate axes which means that two of the three partial derivatives vanish simultaneously and we have the three systems

$$
\begin{aligned}
&F(q_1, q_2, q_3) = 0, \quad \frac{\partial F}{\partial q_1} = 0, \quad \frac{\partial F}{\partial q_2} = 0, \\
&F(q_1, q_2, q_3) = 0, \quad \frac{\partial F}{\partial q_1} = 0, \quad \frac{\partial F}{\partial q_3} = 0, \\
&F(q_1, q_2, q_3) = 0, \quad \frac{\partial F}{\partial q_2} = 0, \quad \frac{\partial F}{\partial q_3} = 0,
\end{aligned}
$$

in three unknowns.

For $l > 3$ parameters the number of polynomials and systems of equations obviously depends on the number of subpolytopes of an l-dimensional cube. The derivation of the polynomials and systems of equations for arbitrary l is straightforward. But now occurs what is called the combinatorial explosion. The number of systems grow exponentially with the number of parameters.

Using other norms ($p = 2$ or $p = \infty$) gives similar equations. In any case we have always equations which are linear combinations of F and their partial derivatives. For $p = 2$ the number of systems is smaller (no subpolytopes), but the equations are more complicated In that case at the intersection point the vector $\mathbf{q}$ must be parallel to the gradient of F, i.e. we have only one system of equations:

$$F(\mathbf{q}) = 0, \quad \frac{\partial F}{\partial \mathbf{q}} = \lambda \mathbf{q}.$$

4. *Solving a system of algebraic equations*

Finding all solutions of a system of algebraic equations is split into two parts. First the algorithm of Buchberger (1985) constructs a so-called Gröbner base. Starting from the equations $F_1(\mathbf{q}) = 0$, $F_2(\mathbf{q}) = 0, \ldots, F_l(\mathbf{q}) = 0$ another system of equations is derived, which has the same set of solutions as the original one. For a zero-dimensional solution set (that means that the system has a finite number of solutions) this may look like

$$
\begin{aligned}
F_1(\mathbf{q}) &= 0 \qquad & G_1(q_1) &= 0, \\
F_2(\mathbf{q}) &= 0 \qquad & G_2(q_1, q_2) &= 0, \\
&\vdots \quad \Leftrightarrow & &\vdots \\
F_l(\mathbf{q}) &= 0 \qquad & G_l(q_1, q_2, \ldots, q_l) &= 0.
\end{aligned}
$$

This new system has a similar form like the triangular form in the case of a system of linear equations. This algorithm is implemented in algebraic and symbolic computation packages like REDUCE or MATHEMATICA. Unfortunately there is until now a severe restriction. Because of time (and storage) limitations it is only possible to attack problems with, at most, five or six parameters.

The second step is the numerical solution of the transformed "tridiagonal" system. Finding the roots of the first equation (a polynomial) gives the first component of the solution vectors. Substituting this value into the second equation yields again a polynomial in one variable. Thus the complete solution vector is obtained. It is not necessary to always find the complete solution vectors. We can omit for example the non-real solutions. Additionally if one or more components would lead to a norm greater than the norm of already known solutions then this candidate cannot lead to the stability radius and its computation can be stopped.

5. *Example*

Given the family of polynomials

$$P(s, q_1, q_2, q_3) = s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0,$$

with

$$
\begin{aligned}
a_3 &= 20 + 2q_2 + 3q_3, \\
a_2 &= 124 + 32q_2 + 42q_3 + 6q_2 q_3, \\
a_1 &= 1040 + 800q_1 + 120q_2 + 120q_3 + 60q_2 q_3, \\
a_0 &= 1600 + 1600q_1.
\end{aligned}
$$

The bounding surfaces are

$$a_0(q_1, q_2, q_3) = 1600 + 1600q_1 = 0,$$

and

$$H_3(q_1, q_2, q_3) = \det$$

$$= 27q_2^2 q_3^3 + \cdots + 21{,}440 = 0.$$

The Hurwitz determinant, not reported explicitly for space limitations, has 24 terms. The stability radius with respect to the real root boundary is $\rho_0 = 1$. The complex root boundary yields $\rho = 0.342$ with the critical frequency $\omega^* = 8.23$. This value is a root of the polynomial

$$H_3(q, -q, -q) = 0.$$

The intersection point lies in a vertex of the cube. For comparison also the results for $p = 2$ and $p = 1$ are given. For these norms the radius with respect to the real root boundary is also $\rho_0 = 1$. For the Euclidian norm the radius with respect to the complex root boundary is $\rho = 0.584$ with the parameters $q_1 = 0.383$, $q_2 = -0.255$ and $q_3 = -0.358$ at the frequency $\omega^* = 8.33$. Using the $\|\ \|_1$-norm the critical system is

$$
\begin{aligned}
&H_3(q_1, 0, q_3) = 0, \\
&\left.\frac{\partial H_3}{\partial q_1}\right|_{q_2=0} + \left.\frac{\partial H_3}{\partial q_3}\right|_{q_2=0} = 0,
\end{aligned}
$$

and the solution is $q_1 = 0.545$, $q_2 = 0$, $q_3 = -0.371$ with $\rho = 0.916$ and $\omega^* = 8.71$. The intersection point lies on an edge of the hexaeder.

6. *Conclusion*

We have shown that in case of polynomial dependency the stability radius can be found by solving systems of algebraic equations. These equations are derived from the bounding surfaces. The difficult problem of finding all solutions of systems is carried out with symbolic and numeric methods. The restriction to few parameters seems to be severe, but a quotation of Lazard (1991) on software for the solution of algebraic equations gives hope for the future. "Five years ago, problems with four or five unknowns were outside of the capabilities of most available softwares. Recent progresses make or will make accessible problems with six or seven unknowns."

References

Barmish, B. R. (1989). A generalization of Kharitonov's four polynomial concept for robust stability problems with linearly dependent coefficient perturbations. *IEEE Trans. Aut. Control*, **34**, 157–165.

Buchberger, B. (1985). Gröbner Bases: an algorithmic method in polynomial ideal theory. *Recent Trends in Multidimensional System Theory*. Bose Ed., Reidel, Dordrecht.

Chapellat, H. and S. P. Bhattacharyya (1988). Exact calculation of stability margin with respect to transfer function coefficients. TCSP Research Report No. 88-003, TX A & M University.

Kaesbauer, D. and J. Ackermann (1990). The distance from stability or Γ-stability boundaries. *Proc. IFAC Congress*, Vol. 5, pp. 130–134.

Lazard, D. (1991). System of algebraic equations (algorithms and complexity). Tutorial paper of the international symposium on symbolic and algebraic computation, Bonn.

Murdoch, T. M. and W. E. Schmittendorf (1991). Use of a genetic algorithm to analyze robust stability problems. Preprint.

Polyak, P. T. and Ya.Z. Tsypkin (1990). Frequency criteria of robust stability and aperiodicity of linear systems. English translation from *Avtomatika i Telemekhanika*, No. 9, pp. 45–54. Plenum, New York.

Vicino, A., A. Tesi and M. Milanese (1990). Computation of nonconservative stability perturbation bounds for systems with nonlinearly correlated uncertainties *IEEE Trans. Aut. Control*, **35**, 835–841.

Automatica, Vol. 29, No 1, pp 219–224, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

μ–K Iteration: A New Algorithm for μ-synthesis*

J.-L. LIN,† I. POSTLETHWAITE† and D.-W. GU†

Key Words—Robust performance; μ-synthesis; D–K iteration; μ–K iteration.

Abstract—In this paper we present a new algorithm for solving robust performance problems by μ-synthesis. Convergence properties of the new algorithm are considered and demonstrated by examples. The algorithm called μ–K iteration, works by flattening the structured singular value μ over frequency.

1. *Introduction*

ROBUST PERFORMANCE is said to be achieved if the design specifications of a controlled system are satisfied in the presence of disturbance signals and model uncertainties. A general framework for analysing robust performance using the structured singular value μ as a measure of performance was introduced by Doyle (1982) who later proposed a controller synthesis procedure, called μ-synthesis. In this procedure a controller is sought which minimizes μ, or which achieves level of performance arbitrarily close to the optimum μ. This minimization problem has not yet been solved, but in Doyle (1985) an approximate solution is given involving a sequence of minimizations, called D–K iteration. The purpose of this paper is to introduce an alternative to D–K iteration, called μ–K iteration.

The robust performance problem under consideration can be described with the help of the now standard feedback configuration shown in Fig. 1.

P is the standard plant which defines the interconnection between the actual plant and the controller, the locations of pertubations representing uncertainty, and the performance function to be minimized; it also includes weighting functions used to describe the uncertainty models and the performance specifications.

Δ is an $n \times n$ block diagonal matrix of perturbations representing uncertainty except for one block which is used to characterize performance. Mathematically Δ is an element of the set

$$\mathbf{\Delta} = \{\mathrm{diag}\,[\delta_1 I_{r_1}, \ldots, \delta_s I_{r_s}, \Delta_1, \ldots, \Delta_f] : \delta_i \in \mathcal{C}, \Delta_j \in \mathcal{C}^{m_j \times m_j}\}, \quad (1)$$

where s is the number of repeated scalar blocks, f is the number of full blocks, and

$$\sum^{s} r_i + \sum^{f} m_j = n. \quad (2)$$

If, as is usual, $\bar{\sigma}(\Delta) \le 1$, then Δ is an element of the bounded set

$$\mathbf{B\Delta} = \{\Delta \in \mathbf{\Delta} : \bar{\sigma}(\Delta) \le 1\}. \quad (3)$$

K is the controller to be designed, and M is the interconnection matrix between P and K given by the linear fractional transformation

$$M = F_l(P, K) := P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}, \quad (4)$$

with P partitioned according to its inputs and outputs as shown in Fig. 1.

The structured singular value of M with respect to $\mathbf{\Delta}$ is defined as

$$\mu_{\mathbf{\Delta}}(M) := \frac{1}{\min\{\bar{\sigma}(\Delta) : \Delta \in \mathbf{\Delta}, \det(I - M\Delta) = 0\}}$$
$$:= 0 \quad \text{if } \det(I - M\Delta) \ne 0, \quad \forall \Delta \in \mathbf{\Delta} \quad (5)$$

The robust performance design problem (μ-synthesis) can now be stated as

$$\inf_{K \text{ stabilizing}} \sup_{\omega \in \mathfrak{R}} \mu_{\mathbf{\Delta}}(M), \quad (6)$$

where $\mathfrak{R}$ denotes the set of real number. This problem has proved difficult to solve and a solution is still not available. However, an approximate solution has been given by Doyle (1985) based on the following bound

$$\mu_{\mathbf{\Delta}}(M) \le \inf_{D \in \mathbf{D}} \bar{\sigma}(DMD^{-1}) \quad (7)$$

where the set $\mathbf{D}$ is defined by

$$\mathbf{D} = \{\mathrm{diag}\,[D_1, \ldots, D_s, d_1 I_{m_1}, \ldots, d_f I_{m_f}] : D_i \in \mathcal{C}^{r_i \times r_i}, D_i = D_i^* > 0, \quad d_j > 0\}. \quad (8)$$

The idea is to look for a solution to

$$\inf_{K \text{ stabilizing}} \sup_{\omega \in \mathfrak{R}} \inf_{D \in \mathbf{D}} \bar{\sigma}[DMD^{-1}], \quad (9)$$

even though the upper bound is not always equal to μ. An approximate method for doing this is to solve (9) first for K keeping D constant, and then for D keeping K constant, and so on. For a fixed D (9) is an $\mathcal{H}^\infty$ optimization problem, and can be solved by various methods. For fixed K (9) can be solved at each frequency by solving a convex optimization problem in D. By taking a sufficient number of frequencies over a sufficiently wide frequency range the scaling matrix D can be approximated by a stable matrix of real rational functions with a stable inverse. This sequence of minimizations, known as D–K iteration, is not guaranteed to converge to the minimum, but it nevertheless offers a systematic procedure for addressing the important problem of robust performance.

The new procedure, μ–K iteration, which will be presented here is motivated by the following.

- In Helton (1985), it is stated that many optimization problems have the property that an optimum solution must make the objective function constant in ω almost everywhere.
- In many examples, using D–K iteration it can be observed that the "μ-optimal" controller appears to flatten $\mu_{\mathbf{\Delta}}(M)$ at least over the bandwidth of the system; constraints on M, for example at high frequencies, can cause a change in the general level of μ.

* Received 4 September 1991; received in final form 1 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor B. A. Francis.

† Control Systems Research, Department of Engineering, Leicester University, Leicester LE1 7RH, U.K.

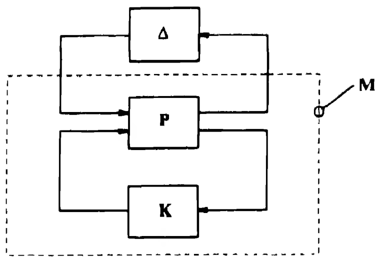


FIG. 1. Standard feedback configuration.

The idea then in the new algorithm is to determine a sequence of controllers which yield a flat structured singular value. This after all is what happens in $\mathscr{H}^\infty$ optimization where the $\mathscr{H}^\infty$-optimal controller results in a cost function with a flat maximum singular value.

To get some insight into how this might be done we will consider a specific robust performance problem for a single-input-single-output plant. This is covered in Section 2. Then in Section 3 the μ-K iteration algorithm is presented for the general robust performance problem. Convergence of the algorithm is considered in Section 4, and two illustrative examples are described in Section 5. Conclusions are given in Section 6.

2. *A SISO robust performance problem*

Consider the control system configuration of Fig. 2, with the following nomenclature

G_0 is the nominal plant, with multiplicative input uncertainty; Δ_2 is the normalized model error, $\|\Delta_2\|_\infty \leq 1$; W_2 is the (model) error bounding function; W_1 is the performance weighting function; Δ_1 is the normalized fictitious uncertainty to characterize performance, $\|\Delta_1\|_\infty \leq 1$.

The configuration can be rearranged into the standard M-Δ structure of Fig. 1, by setting

$$\Delta = \mathrm{diag}\,(\Delta_2, \Delta_1) \tag{10}$$

In which case, the interconnection matrix M is given by

$$M = \begin{bmatrix} -W_2T_0 & -W_2T_0G_0^{-1} \\ W_1S_0G_0 & W_1S_0 \end{bmatrix}, \tag{11}$$

where

$$T_0 := KG_0(I + KG_0)^{-1}, \tag{12}$$

is the nominal complementary sensitivity function, and

$$S_0 := (I + G_0K)^{-1}, \tag{13}$$

is the nominal sensitivity function.

The robust performance problem is to find a stabilizing controller K such that the $\mathscr{H}^\infty$ norm of the transfer function from d to e is less than 1 for all perturbations Δ_2, $\|\Delta_2\|_\infty \leq 1$. This is equivalent to finding a stabilizing controller K such that $\mu_\Delta(M) < 1$, and therefore it makes sense to try to solve (6).

For this relatively simple interconnection matrix M the following facts can be shown.

Fact 1. $\mu_\Delta(M) \leq \bar{\sigma}(M), \quad \forall \omega \in \mathscr{R}.$ (14)

Fact 2. $\bar{\sigma}(M) = \|M\|_F$, the Frobenius norm of M. (15)

Fact 3. $\mu_\Delta(M) = |W_1S_0| + |W_2T_0|$,
where $|\cdot|$ denotes modulus. (16)

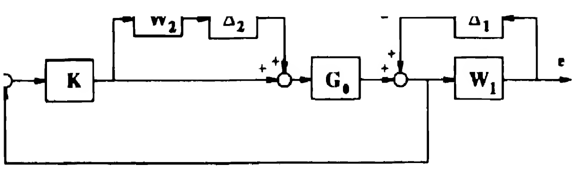


FIG. 2. A SISO robust performance problem.

Fact 4. If at some frequency ω_0

$$|W_1S_0G_0| = |W_2T_0G_0^{-1}| \tag{17}$$

then

$$\mu_\Delta(M) = \bar{\sigma}(M), \tag{18}$$

at the same frequency ω_0.

Suppose now that an $\mathscr{H}^\infty$-optimal controller K_0 is found for M, i.e. we solve

$$\inf_{K \text{ stabilizing}} \|M(K)\|_\infty. \tag{19}$$

It is well known that $\bar{\sigma}[M(K_0)]$ is flat over frequency, and from the above facts (and our observations) $\mu_\Delta[M(K_0)]$ will often have a bandpass-like characteristic as illustrated in Fig. 3. A little thought suggests that a controller which forces $\mu_\Delta(M)$ to be flat will result in a convex $\bar{\sigma}(M)$. Suppose then that we multiply M by a bandpass-like rational function $r(s)$ similar to the shape of $\mu_\Delta[M(K_0)]$ and calculate the $\mathscr{H}^\infty$ optimal controller K_1 for the product rM. One might then expect $\bar{\sigma}[M(K_1)]$ to be convex with $\mu_\Delta[M(K_1)]$ flatter than $\mu_\Delta[M(K_0)]$. This leads us into the μ-K iteration algorithm presented for a general multivariable problem in the next section.

3. *μ-K iteration*

The above discussion motivates the algorithm now proposed for finding an approximate solution to the general robust performance problem

$$\inf_{K \text{ stabilizing}} \sup_{\omega \in \mathscr{R}} \mu_\Delta[F_l(P, K)]. \tag{20}$$

The basic strategy is to "flatten" the μ-curve.

Step 1.

Find the stabilizing $\mathscr{H}^\infty$-optimal controller (a variety of methods exist)

$$K_0 := \arg\inf_K \|F_l(P, K)\|_\infty. \tag{21}$$

The optimization is over stabilizing K, but for notational convenience in (21) and much of what follows the word "stabilizing" has been omitted.

Step 2.

Find the μ-curve corresponding to K_0 (the Matlab toolbox, μ-Tools, could be used for this)

$$\mu_0(j\omega) := \mu_\Delta[F_l(P, K_0)], \tag{22}$$

over a suitable range of frequencies.

Step 3

Normalize $\mu_0(j\omega)$ by its maximum value, i.e. determine

$$\tilde{\mu}_0(j\omega) := \frac{\mu_0(j\omega)}{\|\mu_0(j\omega)\|_\infty}. \tag{23}$$

Step 4.

Find a scalar stable minimum phase real rational function $\tilde{\mu}_0(s)$ by fitting to the $\tilde{\mu}_0(j\omega)$-curve obtained in Step 3.

Step 5.

Multiply the interconnection matrix $F_l(P, K)$ by $\tilde{\mu}_0(s)$. In the specific example of Section 2 this would correspond to multiplying each of the weights W_1 and W_2 by $\tilde{\mu}_0(s)$.

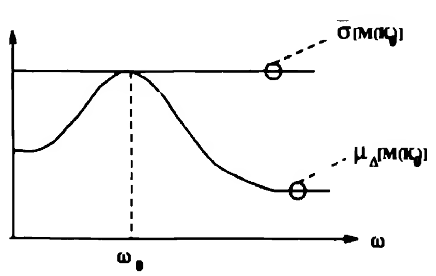


FIG. 3. The maximum and structured singular values of $M(K_0)$.

Step 6
Find the $\mathcal{H}^\infty$-optimal controller

$$K_1(s) = \arg\inf_K \|\mu_0(s)F_l(P, K)\|_\infty \tag{24}$$

Step 7
Find the μ-curve corresponding to K_1

$$\mu_1(j\omega) = \mu_\Delta[F_l(P, K_1)] \tag{25}$$

over the frequency range of interest
Step 8
Normalize $\mu_1(j\omega)$ and denote the result by $\bar{\mu}_1(j\omega)$
Step 9
Curve fit $\bar{\mu}_1(j\omega)$ to get $\mu_1(s)$
Step 10
Find the $\mathcal{H}^\infty$-optimal controller

$$K_2(s) = \arg\inf_K \|\mu_1(s)\mu_0(s)F_l(P, K)\|_\infty \tag{26}$$

Step 11
Find the μ-curve corresponding to K_2

$$\mu_2(j\omega) = \mu_\Delta[F_l(P, K_2)] \tag{27}$$

Subsequent steps of the algorithm should now be clear and in practice would be continued until the μ curve was sufficiently flat over the frequency range of interest or until the desired level of performance (as measured by the peak value of μ) had been reached

4 *Convergence*

In this section we consider the convergence properties of the proposed μ–K algorithm The algorithm generates the following sequences

$$\begin{aligned}
K_0 &= \arg\inf_K \|F_l(P, K)\|_\infty & \mu_0 &= \mu_\Delta[F_l(P, K_0)] \\
K_1 &= \arg\inf_K \|\mu_0 F_l(P, K)\|_\infty & \mu_1 &= \mu_\Delta[F_l(P, K_1)] \\
K_2 &= \arg\inf_K \|\bar{\mu}_1\bar{\mu}_0 F_l(P, K)\|_\infty & \mu_2 &= \mu_\Delta[F_l(P, K_2)]
\end{aligned} \tag{28}$$

Suppose that we normalize each of the μ functions by dividing each curve by its maximum value

$$\bar{\mu}_n = \frac{\mu_n}{\|\mu_n\|_\infty}, \quad n = 0, 1, 2 \tag{29}$$

Then, it is easy to see that

$$0 \le \bar{\mu}_n(j\omega) \le 1, \ \forall\omega \in \mathfrak{R} \quad \text{and} \quad \|\bar{\mu}_n\|_\infty = 1 \quad n = 0, 1, 2 \tag{30}$$

Consider the infinite sequence $\{c_n\}_{n=0}^\infty$ defined by

$$\begin{aligned}
c_0 &= \|F_l(P, K_0)\|_\infty = \inf_K \|F_l(P, K)\|_\infty \\
c_1 &= \|\bar{\mu}_0 F_l(P, K_1)\|_\infty = \inf_K \|\mu_0 F_l(P, K)\|_\infty \\
c_2 &= \|\bar{\mu}_1\mu_0 F_l(P, K_2)\|_\infty = \inf_K \|\bar{\mu}_1\mu_0 F_l(P, K)\|_\infty
\end{aligned} \tag{31}$$

Now because $\bar{\sigma}[\bar{\mu}_{n-1} \cdots \bar{\mu}_0 F_l(P, K_n)]$ is constant in ω it follows from (30) that

$$\|\bar{\mu}_{n-1} \cdots \bar{\mu}_0 F_l(P, K_n)\|_\infty = \|\bar{\mu}_n\bar{\mu}_{n-1} \cdots \mu_0 F_l(P, K_n)\|_\infty \tag{32}$$

and hence

$$\begin{aligned}
c_n &= \|\bar{\mu}_{n-1} \cdots \bar{\mu}_0 F_l(P, K_n)\|_\infty, \\
&= \|\bar{\mu}_n\bar{\mu}_{n-1} \cdots \bar{\mu}_0 F_l(P, K_n)\|_\infty \\
&\ge \inf_K \|\bar{\mu}_n\bar{\mu}_{n-1} \cdots \bar{\mu}_0 f_l(P, K)\|_\infty, \\
&= c_{n+1}
\end{aligned} \tag{33}$$

That is

$$c_n \ge c_{n+1} \ge 0 \tag{34}$$

The sequence $\{c_n\}_{n=0}^\infty$ is therefore monotonically decreasing and bounded, and by the Bolzano–Weierstrass theorem (Bartle 1966) it has a limit point That is

$$c_n \to \text{limit point}, \quad \text{as} \quad n \to \infty \tag{35}$$

We now present a reasoned argument for believing that the sequence $\{\mu_n\}$ will converge to a frequency independent function equal to 1

First, a Lemma which follows from Helton (1985)

Lemma Let $J(\ )$ be a 'well posed' cost function, i e it satisfies Helton's assumptions Then if for a given controller K_i $\sigma[J(K_i)]$ is frequency dependent, then there exists another controller K_j such that $\sigma[J(K_j)]$ is frequency independent and $\|J(K_j)\|_\infty < \|J(K_i)\|_\infty$ ■

Next let

$$J_n(K) = \mu_{n-1} \cdots \mu_0 F_l(P, K) \tag{36}$$

and assume that Helton's assumptions are satisfied Then with this notation we have from (33) that

$$c_n = \|J_n(K_n)\|_\infty \ge \|\mu_n J_n(K_n)\|_\infty \tag{37}$$

Therefore if μ_n is frequency dependent we have by the Lemma that

$$\inf_K \|\mu_n J_n(K)\|_\infty < \|\mu_n J_n(K_n)\|_\infty \tag{38}$$

or equivalently

$$c_{n+1} < c_n \tag{39}$$

But we have already shown that the sequence $\{c_n\}$ converges and therefore the sequence $\{\mu_n\}$ must also converge to a frequency independent function (which must be one by normalization) otherwise $\{c_n\}$ may well decrease below the positive limit

The above argument is clearly lacking in rigour but it does offer support to the observed effectiveness of the algorithm One can also imagine the approach leading to a local minimum where μ is flat but where a controller exists which achieves a lower μ curve which may or may not be flat Like D–K iteration therefore μ–K iteration is not guaranteed to converge to the minimum If it were suspected that μ–K iteration had not given the minimum then it might be possible to get a better result after first rescaling M using a D matrix

5 *Examples*

Two examples are given to illustrate the application of the μ–K iteration algorithm Example 1 is SISO Example 2 is MIMO

Example 1 In this example we solve the following robust performance problem

$$\inf_{K \text{ stabilizing}} \sup_{\omega \in \mathfrak{R}} \mu_\Delta(M)$$

where

$$M = \begin{bmatrix} -W_2T_0G_0^{-1} & -W_1W_2T_0G_0^{-1} \\ S_0 & W_1S_0 \end{bmatrix}$$

$$\mu_\Delta(M) = |W_1S_0| + |W_2T_0G_0^{-1}|$$

and

$$G_0(s) = \frac{0.5(1-s)}{(s+2)(s+0.5)}$$

$$W_1(s) = 50\frac{1 + \dfrac{s}{1.245}}{1 + \dfrac{s}{0.007}}$$

$$W_2(s) = 0.1256\frac{1 + \dfrac{s}{0.502}}{1 + \dfrac{s}{2}}$$

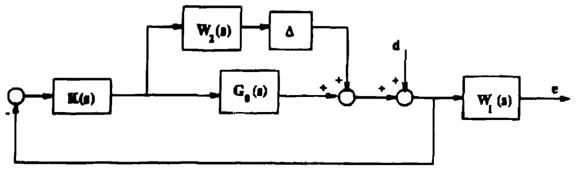


FIG. 4. System with additive uncertainty

The design problem corresponds to meeting disturbance rejection requirements in the presence of plant uncertainty modelled by an additive perturbation; see Fig. 4

Bode magnitude diagrams of the weighting functions and the open-loop gain are shown in Fig. 5. The μ-curve is approximately flat after three μ-K iterations as shown in Fig 6. The controller sequence (after model reduction using balanced truncation) was

$$K_0(s) = 1.923\frac{(s+2)(s+0.5)}{(s+0.007)(s+2.723)},$$

$$K_1(s) = 2.032\frac{(s+2.053)(s+0.498)}{(s+0.007)(s+3.175)},$$

$$K_2(s) = 2.031\frac{(s+2.114)(s+0.5)}{(s+0.007)(s+3.293)}$$

The Bode magnitude diagram of $K_2(s)$ is shown in Fig. 7 The order of the fits used to model $\tilde{\mu}_0(s)$ and $\tilde{\mu}_1(s)$ were ´ and 4, respectively

The problem was repeated using D-K iteration, and similar level of accuracy was achieved after just two iterations, with a second order D scaling The "optimal" controller was

$$K_{D\text{-}K}(s) = 2.0338\frac{(s+2.119)(s+0.5)}{(s+0.007)(s+3.307)}.$$

Example 2. This MIMO example is taken from the MATLAB toolbox manual, μ-TOOLS (Balas *et al.*, 1991) where it is used to demonstrate μ-synthesis. The problem is to meet disturbance rejection requirements in the presence of plant uncertainty modelled as a multiplicative perturbatio at the plant input The plant model is known as HIMAT and represents a scaled version of a remotely piloted aircraft The nominal state-space model is

$$A = \begin{bmatrix} -0.0226 & -36.6 & -18.9 & -32.1 \\ 0 & -1.9 & 0.983 & 0 \\ 0.0123 & -11.7 & -2.63 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

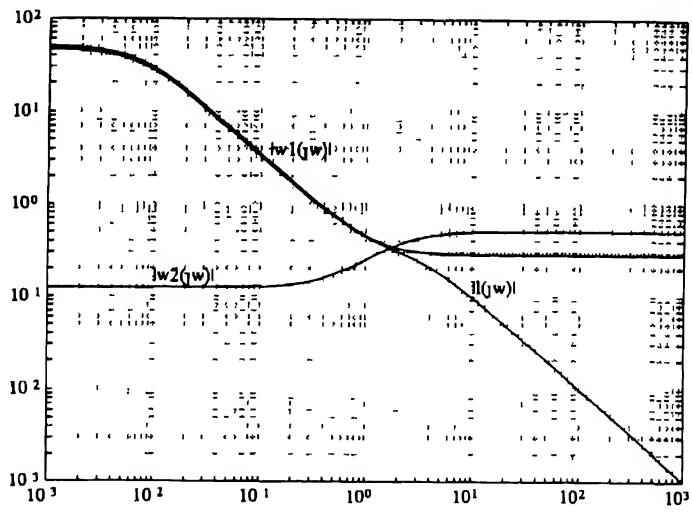


FIG 5 Bode magnitude diagrams of the weighting functions and the open-loop gain (Example 1)

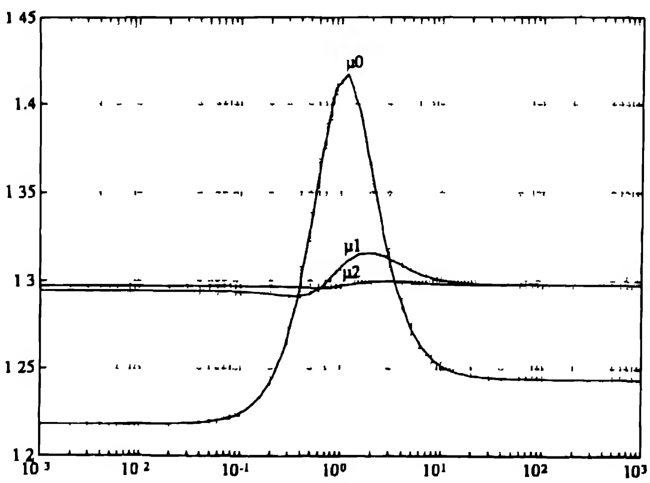


FIG 6 Bode magnitude diagrams of the μ curves of the three μ-K iterations (Example 1)

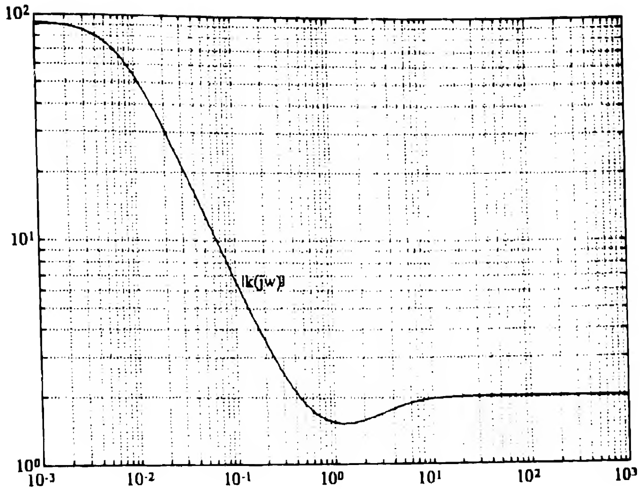


FIG. 7. Bode magnitude diagram of the μ-"optimal" controller (Example 1).

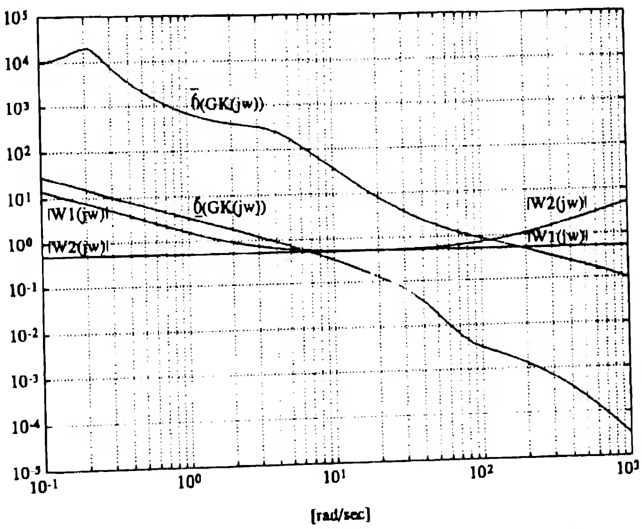


FIG. 8. Bode magnitude diagrams of the weighting functions and the singular values of the open-loop gain (Example 2).

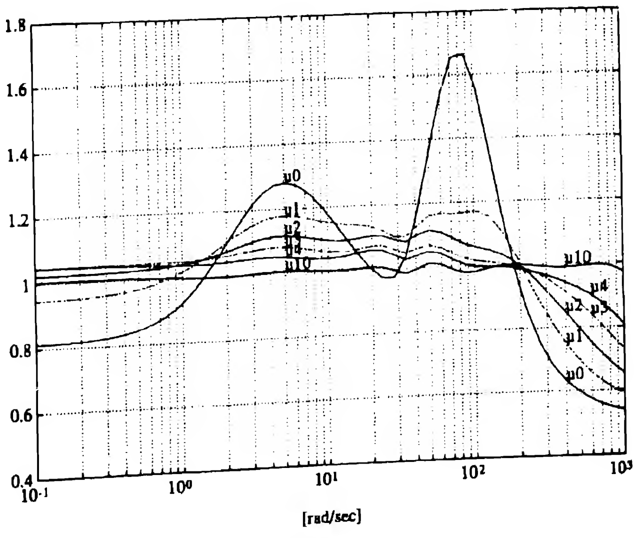


FIG. 9. The μ-curves for several μ–K iterations (Example 2).

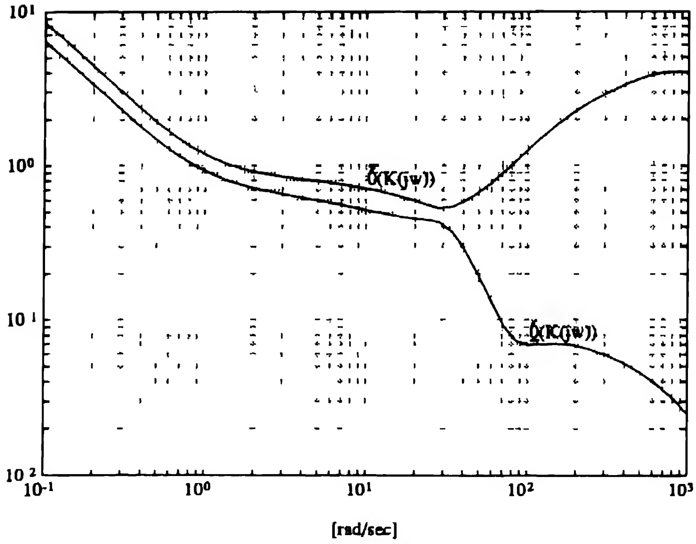


FIG 10 Bode diagrams of the singular values of the μ-"optimal" controller (Example 2)

$$B=\begin{bmatrix} 0 & 0 \\ -0.414 & 0 \\ -77.8 & 22.4 \\ 0 & 0 \end{bmatrix},$$

$$C=\begin{bmatrix} 0 & 57.3 & 0 & 0 \\ 0 & 0 & 0 & 57.3 \end{bmatrix},$$

$$D=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and the weighting functions for the multivariable version of the interconnection matrix M as shown in (11) are

$$W_1(s)=\frac{0.5(s+3)}{s+0.03}I_2,$$

$$W_2(s)=\frac{50(s+100)}{s+10000}I_2,$$

where I_2 is the 2×2 identity matrix

Bode magnitude diagrams of the weighting functions and the singular values of the open-loop gain are shown in Fig 8 The μ-curves for 11 μ–K iterations are shown in Fig 9 The orders of the models used to fit the $\bar{\mu}$-curves range from 2 to 9, and the final controller was adequately represented by a 10th order system The Bode diagrams of the singular values of this controller are shown in Fig 10

A 13th order controller of similar performance was obtained for the same problem using D–K iteration (and balanced truncation) after just four iterations with all D-scalings of order 3

6 *Conclusions*

A new algorithm, μ–K iteration, has been presented for μ-synthesis The accuracy of the algorithm depends on the curve fitting of the $\bar{\mu}(j\omega)$ curves In the examples tested so far the algorithm compares well with D–K iteration and only requires a single scalar function to be fitted over frequency at each iteration Each iteration does, however, require the calculation of μ over a range of frequencies, and this computation is known to be difficult in general Convergence of D–K iteration was more rapid than μ–K in the examples shown

Acknowledgements—The authors are grateful to the U K Science and Engineering Research Council and the Government of Taiwan for financial support

References

Balas, G J , J C Doyle, K Glover, A Packard and R Smith (1991) *μ-Analysis and Synthesis Toolbox* Math Works, Natick, MA

Bartle, R G (1966) *The Elements of Real Analysis* John Wiley, New York

Doyle, J C (1982) Analysis of control systems with structured uncertainties *IEE Proc , Part D,* **129** 242–250

Doyle, J C (1985) Structured uncertainty in control system design In *Proc 24th IEEE Conf on Decision and Control,* Ft Lauderdale, pp 260–265

Helton, J W (1985) Worst case analysis in the frequency domain the H^∞ approach to control *IEEE Trans Aut Control,* **AC-30,** 1154–1170

Automatica Vol 29 No 1 pp 225–228 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Brief Paper

$\mathscr{H}_2$ Control for Discrete-time Systems Optimality and Robustness*†

P L D PERES‡ and J C GEROMEL‡

Key Words—Discrete-time systems, $\mathscr{H}_2$ optimal control, convex programming, linear optimal control, robustness robust control

Abstract—This paper proposes a new approach to determine $\mathscr{H}_2$ optimal control for discrete-time linear systems, based on convex programming It is shown that all stabilizing state feedback control gains belong to a certain convex set well-defined in a special parameter space The Linear Quadratic Problem can be then formulated as the minimization of a linear objective over a convex set The optimal solution of this convex problem furnishes under certain conditions, the same feedback control gain which is obtained from the classical discrete time Riccati equation solution Furthermore, the method proposed can also handle additional constraints, for instance the ones needed to assure asymptotical stability of discrete time systems under actuators failure Some examples illustrate the theory

1 *Introduction*

THE THEORY OF discrete-time linear systems has been developed, historically, as an extension of previous results concerning continuous-time systems In the 1960s the Linear Quadratic Problem—LQP has been exhaustively studied for both continuous-time and discrete time systems becoming a well-known technique for control design (Anderson and Moore, 1971) The discrete-time optimal state feedback solution can be calculated from the solution of an associated Riccati equation It exhibits many good robust properties as for the continuous-time case However the control gain depends directly on the matrices model This fact is one of the main reasons for the difficulties to generalize the discrete LQP to take into account additional requirements as for instance, parameter uncertainties, output feedback decentralized control or sensors/actuators failure Other difficulties stem from the fact that neither the set of stabilizing controllers nor the objective function are convex

This paper presents a convex approach to solve the discrete-time LQP It is formulated as a $\mathscr{H}_2$ optimal state feedback control problem First, we define a parameter space into which a convex set generates all the stabilizing state feedback control gains An optimization problem is then formulated whose solution furnishes the state feedback gain that minimizes the $\mathscr{H}_2$ norm of a closed-loop transfer function Under certain conditions (to be defined in the sequel) this control gain equals the one provided by the classical LQP approach, that is, the associated Riccati equation solution

Furthermore, additional requirements can be easily incorporated in the algorithm To support this claim, we solve completely the discrete-time linear system stabilizability problem subject to actuators failure In this case, since only one feedback gain must be stabilizing for all prespecified contingencies, we provide an upper bound for each closed-loop transfer function $\mathscr{H}_2$ norm

* Received 17 July 1991, received in final form 18 December 1991 This paper was not presented at any IFAC meeting This paper was recommended for publication in revised form by Guest Editor P Dorato

† This research has been supported in part by grants from "Fundação de Amparo à Pesquisa do Estado de Sao Paulo—FAPESP" and "Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq", Brazil

‡ LAC-DT/Faculty of Electrical Engineering UNICAMP, CP 6101, 13081, Campinas, SP, Brazil

2 *Preliminaries*

Let us consider a discrete time linear system whose dynamic behavior is given by the following difference equations

$$
\begin{aligned}
x_{k+1} &= Ax_k + B_1 w_k + B_2 u_k, \\
u_k &= -Kx_k \\
z_k &= Cx_k + Du_k,
\end{aligned} \tag{1}
$$

where $x_k \in \Re^n$ is the state variable $u_k \in \Re^m$ is the control variable $w_k \in \Re^l$ is the external disturbance and $z_k \in \Re^q$ is the output variable Matrices A, B_1 B_2 C and D have appropriate dimensions and are supposed to be known Without loss of generality, the usual orthogonality hypothesis is also made, that is $C'D = 0$ and $D'D > 0$ Defining the closed-loop matrices $A_{cl} = A - B_2K$ and $C_{cl} = C - DK$ and supposing that a state feedback gain K is calculated in such a way that A_{cl} is asymptotically stable the closed-loop transfer function from w to z is given by

$$H(z) \triangleq C_{cl}[zI - A_{cl}]^{-1}B_1 \tag{2}$$

The $\mathscr{H}_2$ norm for a stable transfer matrix $H(z)$ can be defined as

$$\|H\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Tr}\,\{H(e^{-j\omega})'\, H(e^{j\omega})\}\, d\omega \tag{3}$$

and can also be calculated from the discrete-time associated Gramians Let L_c be the controllability Gramian of (A_{cl}, B_1) and L_o the observability Gramian of (C_{cl}, A_{cl}) Then,

$$A_{cl}L_cA_{cl}' - L_c + B_1B_1' = 0 \tag{4}$$

$$A_{cl}'L_oA_{cl} - L_o + C_{cl}'C_{cl} = 0 \tag{5}$$

The $\mathscr{H}_2$ norm is given by

$$\|H\|_2^2 = \mathrm{Tr}\,(C_{cl}L_cC_{cl}') = \mathrm{Tr}\,(B_1'L_oB_1) \tag{6}$$

If we denote by $\mathscr{K}$ the set of all stabilizing state feedback control gains $K \in \Re^{m\times n}$ the problem of $\mathscr{H}_2$ optimal control can be stated as follows

$$\text{(P1)} \qquad \min\,\{\|H\|_2^2 : K \in \mathscr{K}\} \tag{7}$$

Examined in the state feedback parameter space, this problem may have a very complicated geometry In fact, neither the objective function nor $\mathscr{K}$ are convex with respect to the elements of the control gain K However, it is well known (Anderson and Moore, 1971, Dorato and Levis, 1971) that the optimal solution of (P1) is given by

$$K = (B_2'PB_2 + D'D)^{-1}B_2'PA, \tag{8}$$

where, under controllability and observability assumptions, $P \in \Re^{n\times n}$ is the unique symmetric positive definite solution of the discrete-time Riccati equation

$$A'PA - P - A'PB_2(B_2'PB_2 + D'D)^{-1}B_2'PA + C'C = 0. \quad (9)$$

From (8), we note that the optimal feedback gain K depends directly on the system matrices A and B_2. Obviously (Kwakernaak and Sivan, 1972) this solution can also be regarded as the optimal one for the following discrete-time LQP:

$$\text{(P2)} \qquad \min_{u_k} \sum_{k=0}^{\infty} (x_k'C'Cx_k + u_k'D'Du_k), \qquad (10)$$
$$x_{k+1} = Ax_k + B_2u_k, \quad x_0 \text{ given}.$$

It is important to remark that, although the optimal control gain does not depend on the initial condition x_0, the minimal value of the quadratic criterion is such that

$$J^* = x_0'Px_0. \quad (11)$$

If we rearrange equation (9) in a closed-loop form, keeping in mind that $C'D = 0$, we have

$$(A - B_2K)'P(A - B_2K) - P + (C - DK)'(C - DK) = 0, \quad (12)$$

and we conclude from (5) that $P = L_o$ and $\|H\|_2^2 = J^*$ provided we choose $B_1 = x_0$. Note that the necessary optimality conditions for (P1) are given by equations (4)–(5) and

$$[(B_2'L_oB_2 + D'D)K - B_2'L_oA]L_c = 0, \quad (13)$$

being thus related to the solution of (P2). Indeed, from (13), if L_c is non singular then the unique solution is given by (8), which is the optimal solution for (P2) and does not depend on the initial condition x_0. This fact will always occur in case (A_{cl}, B_1') is observable or rank $(B_1) = n$ (a sufficient condition). This is assumed throughout this paper.

3. *Main results*

In this section, we propose a new convex programming problem which is equivalent to (P1) in a sense to be defined in the sequel. From the above discussion, this is necessary in order to circumvent the non convexity of $\mathcal{K}$. First of all, let us introduce the following extended matrices $F \in \Re^{p\times p}$, $p \triangleq m + n$ and $G \in \Re^{p\times m}$

$$F = \begin{bmatrix} A & -B_2 \\ 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad (14)$$

as well as the symmetric matrices $Q \in \Re^{p\times p}$, $R \in \Re^{p\times p}$

$$Q = \begin{bmatrix} B_1B_1' & 0 \\ 0 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} C'C & 0 \\ 0 & D'D \end{bmatrix}, \quad (15)$$

and the set $\mathscr{C}_2$:

$$\mathscr{C}_2 \triangleq \{\mathcal{W} = \mathcal{W}' \geq 0 : v'[F\mathcal{W}F' - \mathcal{W} + Q]v \leq 0, \\ \forall v \in \Re^p \neq 0 : G'v = 0\}, \quad (16)$$

where $\mathcal{W} \in \Re^{p\times p}$ is symmetric and is partitioned as

$$\mathcal{W} = \begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix}, \quad (17)$$

with $W_1 \in \Re^{n\times n}$ being positive definite.

Theorem 1. The following statements hold:
(a) $\mathscr{C}_2$ is a convex set.
(b) The pair (A, B_2) is stabilizable by a linear state feedback if and only if $\mathscr{C}_2 \neq \varnothing$. In the affirmative case, $\forall \mathcal{W} \in \mathscr{C}_2$, a stabilizing control gain is given by $K = W_2'W_1^{-1}$.
(c) For a $\mathcal{W}^* \notin \mathscr{C}_2$, there exists an hyperplane that separates $\mathcal{W}^*$ from $\mathscr{C}_2$.

Proof. Part (a) follows directly from the fact that $\mathscr{C}_2$ is defined as the intersection of an uncountable number of linear constraints. Let us prove part (b). Suppose the pair (A, B_2) is stabilizable; then there exists a control gain $K \in \Re^{m\times n}$ as well as a symmetric positive definite matrix $L_c \in \Re^{n\times n}$, $L_c > 0$, satisfying

$$A_{cl}L_cA_{cl}' - L_c = -B_1B_1'. \quad (18)$$

As a consequence, there also exists $W \geq L_c$ such that

$$A_{cl}WA_{cl}' - W + B_1B_1' \leq 0. \quad (19)$$

Keeping in mind that $A_{cl} = A - B_2K$ and developing (19), it gives, for any $x \in \Re^n$

$$x'[AWA' - W - B_2KWA' - AWK'B_2' \\ + B_2KWK'B_2' + B_1B_1']x \leq 0, \quad (20)$$
$$v' = [x' \vdots 0].$$

Since $\forall v \neq 0 \in \Re^p : G'v = 0$, $v = [x' \vdots 0]$, it is easy to verify from (20) that the $\mathcal{W}$ matrix

$$\mathcal{W} = \begin{bmatrix} W & WK' \\ KW & KWK' \end{bmatrix} \geq 0, \quad (21)$$

belongs to $\mathscr{C}_2$, proving thus the necessity.

Now, the sufficiency. From the fact that every $v \neq 0 \in \Re^p : G'v = 0$ has the form $v' = [x' \vdots 0]$, with $x \neq 0 \in \Re^n$, taking $\mathcal{W} \in \mathscr{C}_2$ partitioned as in (17), we have

$$\begin{aligned} & v'[F\mathcal{W}F' - \mathcal{W} + Q]v \\ &= x'[AW_1A' - W_1 - B_2W_2'A' \\ &\quad - AW_2B_2' + B_2W_3B_2' + B_1B_1']x \\ &= x'[(A - B_2W_2'W_1^{-1})W_1(A - B_2W_2'W_1^{-1})' \\ &\quad - W_1 + B_1B_1']x \\ &\quad + x'[B_2(W_3 - W_2'W_1^{-1}W_2)B_2']x. \end{aligned} \quad (22)$$

Since $\mathcal{W} \in \mathscr{C}_2$, we have

$$v'[F\mathcal{W}F' - \mathcal{W} + Q]v \leq 0 \quad \forall v \in \Re^p \neq 0 : G'v = 0 \\ \mathcal{W} \geq 0 \Rightarrow W_3 \geq W_2'W_1^{-1}W_2. \quad (23)$$

Consequently, from (22) and (23), we conclude that

$$x'[(A - B_2W_2'W_1^{-1})W_1(A - B_2W_2'W_1^{-1})' - W_1 + B_1B_1']x \leq 0. \quad (24)$$

The last inequality evidences that the control gain $K = W_2'W_1^{-1}$ is a stabilizing one for the pair (A, B_2). The proof of part (b) then follows. Part (c) is obvious, since $\mathscr{C}_2$ is a convex set (Luenberger, 1973).

At this point, it is important to notice, from the previous theorem, that there exists a one-to-one relationship between the elements of $\mathcal{K}$ (the set of all stabilizing control gains) and the ones of the convex set $\mathscr{C}_2$. Indeed, for each $\mathcal{W} \in \mathscr{C}_2$, $K = W_2'W_1^{-1} \in \mathcal{K}$ and, conversely, for each $K \in \mathcal{K}$, there exists a matrix $\mathcal{W}$ given by (21) which belongs to $\mathscr{C}_2$ (see Fig. 1).

Now, with the results of Theorem 1, it is possible to handle the convex set $\mathscr{C}_2$ in order to achieve stabilization, avoiding the use of set $\mathcal{K}$. The next theorem guarantees the optimal characteristics to a particular $\mathcal{W} \in \mathscr{C}_2$, which furnishes the desired optimal control gain.

Theorem 2. The optimal solution of (P1) is obtained by solving the following convex problem:

$$\text{(P3)} \qquad \min \{\mathrm{Tr}\,(R\mathcal{W}) : \mathcal{W} \in \mathscr{C}_2\}. \quad (25)$$

Being $\mathcal{W}^*$ its optimal solution, then $K^* = W_2'W_1^{-1}$ solves (P1) (and consequently solves also (P2)) and $\mathrm{Tr}\,(R\mathcal{W}^*) = J^*$.

Proof. From Theorem 1, $K^* = W_2'W_1^{-1} \in \mathcal{K}$, being so a

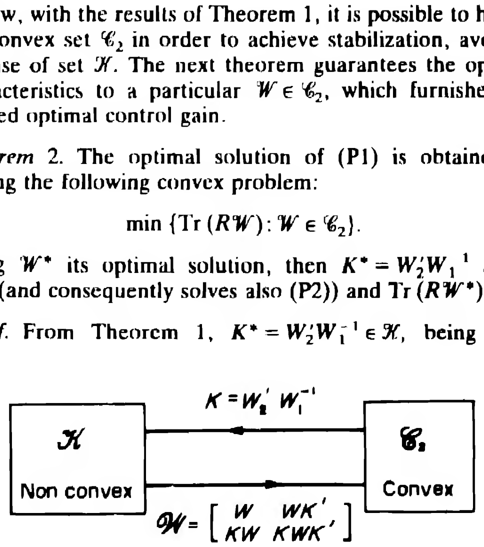


FIG. 1. Relationship between sets $\mathcal{K}$ and $\mathscr{C}_2$.

stabilizing control gain. We just have to prove that this state feedback is indeed the optimal one. First, let us prove that for $K = W_2'W_1^{-1}$, $\forall \mathcal{W} \in \mathscr{C}_2$, we have $\|H\|_2^2 \le \mathrm{Tr}\,(R\mathcal{W})$. To this end, note that for any $\mathcal{W}$ belonging to $\mathscr{C}_2$, W_1 is positive definite and

$$(A - B_2K)W_1(A - B_2K)' - W_1 + B_1B_1' \le 0, \tag{26}$$

then we conclude from (4) that $W_1 \ge L_c$. Now, using equation (6), the fact that $\mathcal{W} \ge 0$ implies $W_3 \ge W_2'W_1^{-1}W_2$ and the orthogonality condition $C'D = 0$, we get

$$\begin{aligned}\|H\|_2^2 &= \mathrm{Tr}\,\{(C - DK)L_c(C - DK)'\}\\ &\le \mathrm{Tr}\,\{(C - DK)W_1(C - DK)'\}\\ &\le \mathrm{Tr}\,\{CW_1C' + DW_2'W_1^{-1}W_2D'\}\\ &\le \mathrm{Tr}\,\{R\mathcal{W}\}, \quad \forall \mathcal{W} \in \mathscr{C}_2.\end{aligned} \tag{27}$$

The above inequality is a very important result; actually it holds in particular for the optimal solution of (P1), implying that $J^* \le \mathrm{Tr}\,(R\mathcal{W})$, $\forall \mathcal{W} \in \mathscr{C}_2$ and, as can be verified, the equality holds for

$$\mathcal{W}^* = \begin{bmatrix} L_c & L_cK' \\ KL_c & KL_cK' \end{bmatrix} \in \mathscr{C}_2, \tag{28}$$

where $K = (B_2'L_oB_2 + D'D)^{-1}B_2'L_oA$, being $P = L_o$ the definite positive solution of the discrete-time Riccati equation (9) and L_c being the solution of the controllability Gramian (equation (4)). Now, it is simple to see that $K^* = W_2'W_1^{-1}$ provides the optimal gain and

$$\begin{aligned}\min\,\{\mathrm{Tr}\,(R\mathcal{W}) : \mathcal{W} \in \mathscr{C}_2\} &= \mathrm{Tr}\,(R\mathcal{W}^*)\\ &= \mathrm{Tr}\,(C_{cl}L_cC_{cl}') = J^*\end{aligned} \tag{29}$$

The proof of Theorem 2 is completed.

The above results deserve some remarks The first one is related to the uniqueness of solution of (P3) Being convex (not strictly), it would be possible that its optimal solution was not unique in the sense that different state feedback gains could be generated Fortunately, this fact does not occur. To prove that, suppose $\mathcal{W}^* \ne \mathcal{W} \in \mathscr{C}_2$ generate two different state feedback gains, being such that $\mathrm{Tr}\,\{R(\mathcal{W} - \mathcal{W}^*)\} = 0$. With K associated to $\mathcal{W}$, determining the transfer function $H(z)$, using (27) and taking into account that (P1) admits only one solution, we obtain

$$\begin{aligned}J^* &= \min\,\{\|H\|_2^2 : K \in \mathscr{K}\},\\ &< \|H\|_2^2 \le \mathrm{Tr}\,(R\mathcal{W}),\end{aligned} \tag{30}$$

which is an impossibility, since by assumption $J^* = \mathrm{Tr}\,(R\mathcal{W}^*)$. The second remark concerns the geometry of (P3). Indeed, we have proved that the discrete-time LQP is equivalent to a convex problem defined on a special parameter space (the elements of matrix $\mathcal{W} \in \mathscr{C}_2$). This fact is of great importance since (P3) can be solved by means of efficient numerical procedures available in the literature (see for instance a convex-based algorithm proposed in Geromel *et al.*, 1991).

Another important characteristic of (P3) is that further convex constraints can be easily added to it For instance, suppose we want to solve the problem of finding a control gain that guarantees closed-loop stability and minimal suboptimality in case of actuator failure. This problem can be stated as follows: suppose the model (1) is such that the input matrix B_2 is not exactly known but belongs to the set $\mathscr{B}_2 \triangleq \{B_{2i} : i = 1 \cdots M\}$. Find (if one exists) a state feedback matrix gain K_f and a positive parameter μ_f (as small as possible) such that:

- $A - B_2K_f$ is asymptotically stable $\forall B_2 \in \mathscr{B}_2$.
- $\|H\|_2^2 \le \mu_f$, $\forall B_2 \in \mathscr{B}_2$.

Note that, in the above problem formulation, we are representing a particular actuator failure (the ith component of the control variable u is reduced to zero) by the equality $B_2u = B_{2i}u$. In this sense, the set $\mathscr{B}_2$ is composed by all matrices obtained from B_2 by zeroing its ith column, $i = 1 \cdots m$. In fact, this is the case when only one actuator is supposed to fail at a time. However, it is obvious that any other situation can be taken into account by simple including the case in $\mathscr{B}_2$ and defining properly the parameter M (see the example in the next section).

Theorem 3. Define the convex set $\mathscr{C}_{2f} = \bigcap_{i=1}^{M} \mathscr{C}_{2i}$, where $\mathscr{C}_{2i}$ is given by (16) with F replaced by F_i (defined as in (14)), $i = 1 \cdots M$. The optimal solution $\mathcal{W}_f$ of

$$\text{(P4)} \qquad \min\,\{\mathrm{Tr}\,(R\mathcal{W}) : \mathcal{W} \in \mathscr{C}_{2f}\}, \tag{31}$$

is such that $K_f = W_2'W_1^{-1}$ and $\mu_f = \mathrm{Tr}\,(R\mathcal{W}_f)$ solve the actuator's failure problem stated before

Proof. Since $\mathcal{W}_f$ belongs to $\mathscr{C}_{2f}$, it also belongs to each one of the sets $\mathscr{C}_{2i}$, $i = 1 \cdots M$; then, we have $\forall v \in \mathbb{R}^p \ne 0$, $G'v = 0$,

$$v'[F_i\mathcal{W}_fF_i' - \mathcal{W}_f + Q]v \le 0, \quad \forall i = 1 \cdots M, \tag{32}$$

implying (see the proof of Theorem 1), with $\mathcal{W}_f$ partitioned as in (17), that

$$(A - B_{2i}W_2'W_1^{-1})W_1(A - B_{2i}W_2'W_1^{-1})' - W_1 + B_1B_1' \le 0, \quad \forall i = 1 \cdots M \tag{33}$$

It is then clear that $K_f = W_2'W_1^{-1}$ guarantees asymptotical stability for all $B_2 \in \mathscr{B}_2$, taking into account all feasible actuator's failure On the other hand, since W_1 is unique for all the above $i = 1 \cdots M$ inequalities, using (27) we get

$$\mathrm{Tr}\,(R\mathcal{W}_f) \triangleq \mu_f \ge \|H\|_2^2, \quad \forall B_{2i}, \quad i = 1 \cdots M, \tag{34}$$

consequently, $\mu_f \ge \|H\|_2^2$, $\forall B_2 \in \mathscr{B}_2$ and the proof is completed. □

4 *Numerical procedure and examples*

This section is in part devoted to solve the convex problem (P4), that is

$$\min\,\{\mathrm{Tr}\,(R\mathcal{W}) : \mathcal{W} \in \mathscr{C}_{2f}\}. \tag{35}$$

Note that this problem is convex and reduces to (P3) in case $M = 1$. Using the previous results, for each $\mathcal{W}_\ell \notin \mathscr{C}_{2f}$, it is a simple task (Geromel *et al.*, 1991) to determine a matrix $\alpha(\mathcal{W}_\ell)$ and a scalar $\beta(\mathcal{W}_\ell)$, which define a separating hyperplane from $\mathcal{W}_\ell$ to $\mathscr{C}_{2f}$. Then, it is possible to apply the following outer linearization algorithm which converges to the global optimal solution (Bernussou *et al.*, 1989; Luenberger, 1973):

$$\mathcal{W}_{\ell+1} = \arg\min\,\{\mathrm{Tr}\,(R\mathcal{W}) : \mathcal{W} \in \mathscr{C}_{2f}^{\ell+1}\}, \tag{36}$$

$$\mathscr{C}_{2f}^{\ell+1} = \mathscr{C}_{2f}^{\ell} \cap \{\mathcal{W} : \beta(\mathcal{W}_\ell) + \langle \alpha(\mathcal{W}_\ell), \mathcal{W} \rangle \le -\varepsilon\}, \tag{37}$$

$$\mathscr{C}_{2f}^{0} \text{ given},$$

where $\varepsilon > 0$ is a sufficiently small parameter used to approximate $\mathscr{C}_{2f}$ by a closed convex set and $\mathscr{C}_{2f}^0 \supset \mathscr{C}_{2f}$.

To illustrate the proposed method, let us consider the following discrete-time system:

$$A = \begin{bmatrix} 0.2113 & 0.0087 & 0.4524 \\ 0.0824 & 0.8096 & 0.8075 \\ 0.7599 & 0.8474 & 0.4832 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0.6135 & 0.6538 \\ 0.2749 & 0.4899 \\ 0.8807 & 0.7741 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and $B_1 = \mathbf{I}$. Since the eigenvalues of matrix A are $(0.3827, -0.4919, 1.6133)$, it is open-loop unstable Using the discrete-time Riccati equation, we get

$$K_{\text{Riccati}} = \begin{bmatrix} 0.2968 & 0.3758 & 0.3114 \\ 0.2302 & 0.4953 & 0.4997 \end{bmatrix},$$

and $\|H\|_2^2 = 5.2448$ With the proposed algorithm, the

optimal solution of (P3) is calculated to be

$$W^* = \begin{bmatrix} 1.4582 & -0.3930 & -0.0926 & 0.2597 & 0.0918 \\ -0.3930 & 1.6316 & -0.1254 & 0.4629 & 0.6612 \\ -0.0926 & -0.1254 & 1.1972 & 0.3013 & 0.5151 \\ 0.2597 & 0.4629 & 0.3013 & 0.3489 & 0.4411 \\ 0.0918 & 0.6612 & 0.5151 & 0.4411 & 0.6084 \end{bmatrix}$$

providing the control gain

$$K^* = \begin{bmatrix} 0.3006 & 0.3803 & 0.3148 \\ 0.2292 & 0.4989 & 0.5002 \end{bmatrix},$$

and $\|H\|_2^2 = 5.2452$. Comparing both numerical solutions, we verify relative errors of about 0.8% in the control gain norm and about 0.007% in the $\mathcal{H}_2$ norm.

Now, let us suppose that one of the actuators may fail. Obviously, the optimal solution of the discrete-time LQP provides no guarantee for stability in this case. However, Theorem 3 can easily handle this possibility by defining $\mathcal{B}_2 \triangleq \{B_{21}, B_{22}, B_{23}\}$ where

$$B_{21} = \begin{bmatrix} 0.6135 & 0.6538 \\ 0.2749 & 0.4899 \\ 0.8807 & 0.7741 \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 0.6135 & 0 \\ 0.2749 & 0 \\ 0.8807 & 0 \end{bmatrix},$$

$$B_{23} = \begin{bmatrix} 0 & 0.6538 \\ 0 & 0.4899 \\ 0 & 0.7741 \end{bmatrix}$$

Considering the set $\mathcal{B}_2$ and solving (P4), we have found $\mu_f = 11.8127$ and the control gain

$$K_f = \begin{bmatrix} 0.2966 & 0.6356 & 0.7405 \\ 0.4823 & 0.6592 & 0.4925 \end{bmatrix},$$

that guarantees the closed-loop asymptotical stability under the actuators failure previously defined. Furthermore, for each matrix B_{2i}, $i = 1, 2, 3$, the associated $\mathcal{H}_2$ norm are given by

$$B_{21} : \|H\|_2^2 = 7.0560,$$
$$B_{22} : \|H\|_2^2 = 8.6285,$$
$$B_{23} : \|H\|_2^2 = 8.4779,$$

making evident that μ_f is a $\mathcal{H}_2$-norm upper bound indeed. Table 1 shows the closed-loop eigenvalues in two situations, namely $B_2 = B_{21}$ (the nominal one) and $B_2 = B_{22}$ (actuator-2 failure). It is easy to see that under the last contingency the closed-loop system with K_{Riccati} becomes unstable, and the same does not occur for the closed-loop system with K_f. In Figs 2 and 3, we show the impulse response (and unitary impulse has been applied to the second component of w) of the closed-loop system with K_{Riccati} and K_f, respectively, supposing that actuator 2 (corresponding to the second component of u) fails at $k = 6$ and remains inactive for $k > 6$. The unstable behavior is obvious when the Riccati control gain is used.

5. *Conclusion*

In this paper we have proved that the optimal state feedback solution of the discrete-time LQP can be determined by means of a convex problem. This is an important result, mainly due to two facts. First, for a given pair (A, B_2), all stabilizing state feedback gains can be parametrized over a convex set. Second, the convexity of the $\mathcal{H}_2$ discrete-time control problem allows to solve it by using the most powerful mathematical programming methods. As a by-product of this fact, we claim that additional structural constraints can be easily handled. For instance, we defined and solved a problem involving actuator's failure whose solution, to our knowledge, was not available until this time in the literature. In this sense we want to emphasize that the numerical procedure proposed here is specially addressed for solving "non-classical" $\mathcal{H}_2$ optimal control problems, including additional convex constraints (actuators/sensors failure, robustness, uncertain systems control, . . .) which cannot be solved by means of the algebraic Riccati equation.

TABLE 1. CLOSED-LOOP EIGENVALUES

Contingencies	K_{Riccati}	K_f
Both actuators	−0.4826	0.4153
	0.2085	−0.5326
	0.4385	−0.4069
Only first actuator	−0.4861	0.4047
	0.3347	0.5973
	1.0959	−0.5068

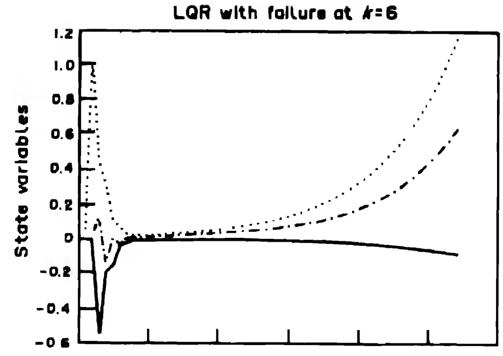


FIG. 2. Impulse response—K_{Riccati}.

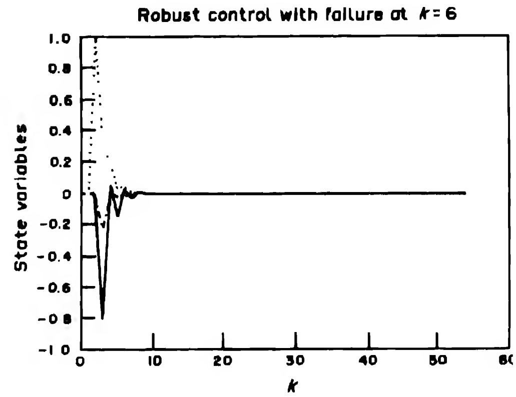


FIG. 3. Impulse response—K_f.

References

Anderson, B. D. O. and J. B. Moore (1971). *Linear Optimal Control.* Prentice Hall, Englewood Cliffs, NJ.

Bernussou, J., P. L. D. Peres and J. C. Geromel (1989). A linear programming oriented procedure for quadratic stabilization of uncertain systems. *Systems and Control Letters*, **13**, 65–72.

Dorato, P. and A. H. Levis (1971). Optimal linear regulators: the discrete time case. *IEEE Trans. on Aut. Control*, **16**, 613–620.

Geromel, J. C., P. L. D. Peres and J. Bernussou (1991). On a convex parameter space method for linear control design of uncertain systems. *SIAM J. on Control and Optimiz.*, **29**, 381–402.

Kwakernaak, H. and R. Sivan (1972). *Linear Optimal Control Systems.* John Wiley, New York.

Luenberger, D. G. (1973). *Introduction to Linear Programming*. Addison-Wesley, Reading, MA.

Automatica, Vol 29, No 1, pp 229 236, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Brief Paper

A Monte Carlo Approach to the Analysis of Control System Robustness*

LAURA RYAN RAY† and ROBERT F. STENGEL‡

Key Words—Robustness; multivariable control systems; control system analysis; Monte Carlo methods.

Abstract—*Stochastic robustness*, a simple technique used to estimate the stability and performance robustness of linear, time-invariant systems, is described. The scalar *probability of instability* is introduced as a measure of stability robustness. Examples are given of stochastic performance robustness measures based on classical time-domain specifications. The relationship between stochastic robustness measures and control system design parameters is discussed. The technique is demonstrated by analysing an LQG/LTR system designed for a flexible robot arm. It is concluded that the analysis of stochastic robustness offers a good alternative to existing robustness metrics. It has direct bearing on engineering objectives, and it is appropriate for evaluating robust control system synthesis methods currently practised

Introduction

STANDARD LINEAR CONTROL system design methods rely on accurate models of the system to be controlled. Because models are never perfect, robustness analysis is necessary to determine the possibility of instability or inadequate performance in the face of uncertainty. Synthesis of robust control system is predicted on a good measure of robustness Consequently, much research activity during the past two decades has been devoted to developing adequate measures of robustness for linear, time-invariant systems that can in turn be used for robust control system synthesis. In most instances, robustness is treated deterministically, using singular-value analysis (e.g. Lehtomaki *et al.*, 1981; Doyle, 1982) or parameter-space methods (e.g. Siljak, 1989, Vicino *et al.*, 1990). These methods can be applied without regard to actual bounds on system parameters; hence, the relationship of the metric to uncertainties in the physical system often is weak. Furthermore, deterministic metrics can be conservative and/or difficult to determine for systems with many uncertain parameters. Consequently, overconservative control system designs or designs that are insufficiently robust in the face of real world uncertainties are a danger.

Stochastic Robustness Analysis (SRA) uses statistical descriptions of parameter uncertainty to determine whether stability/performance robustness criteria are met. Stengel (1980) introduced Monte Carlo analysis of the scalar probability of instability, which is central to the analysis of stability robustness. SRA is described in more detail in Stengel and Ray (1991); exact confidence intervals for the scalar probability of instability are presented, and the stochastic root locus, or probability density of the closed-loop eigenvalues, is shown to portray robustness properties graphically. Ray and Stengel (1991) illustrates the use of SRA to compare control system designs for full-state feedback aircraft control systems and to analyse systems with finite-dimensional uncertain dynamics. Because it provides a statistical measure of robustness and because it uses knowledge of the statistics of parameter variations, SRA is inherently intuitive and accurate. The physical meaning behind the probability of instability is apparent, and overconservative or insufficiently robust designs can be avoided.

Concepts behind stochastic *stability* robustness are readily extended to provide measures of *performance* robustness. Design specifications such as rise time, peak overshoot, settling time, and steady-state error are normally used as indicators of adequate performance and lend themselves to the same kind of analysis as described above. SRA can be applied to classical criteria giving probabilistic bounds on scalar performance measures. Metrics resulting from stability and performance robustness can be related to controller parameters, thus providing a foundation for design tradeoffs and optimization. This paper summarizes stochastic stability and performance robustness analysis The analysis is illustrated by studying the effectiveness of robustness recovery on a stochastic optimal control system with parameter uncertainties.

Stochastic stability robustness

Stochastic stability robustness is described in Stengel and Ray (1991) and is summarized here. Consider a linear, time-invariant system where the dynamic, control, and output matrices, $\mathbf{F(p)}$, $\mathbf{G(p)}$, and $\mathbf{H(p)}$ may be arbitrary functions of an r-dimensional parameter vector, $\mathbf{p}$. The control gain matrix $\mathbf{C}$ is designed using some nominal or "mean" value of the dynamic model, denoted $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, that represents $\mathbf{F(p)}$, $\mathbf{G(p)}$, $\mathbf{H(p)}$ evaluated at the nominal parameter vector. The actual system has an unknown description that depends on the actual (unknown) value of the parameter vector $\mathbf{p}$. Environment, variations in the nominal state, system failures, parameter estimation errors, wear, and manufacturing differences all can contribute to mismatch between the actual system and that used to design the controller. For SRA, $\mathbf{p}$ is assumed to have a known or estimated probability density function, denoted $\mathrm{pr}(\mathbf{p})$, that expresses the parameter uncertainty statistics due to the above factors.

The eigenvalues of the matrix $[\mathbf{F(p)} \quad \mathbf{G(p)CH(p)}]$ determine closed-loop stability. To estimate the *probability of instability* ($\mathbb{P}$) using *Monte Carlo* evaluation, the closed-loop eigenvalues are evaluated J times with each element of $\mathbf{p}$ specified by a random-number generator whose individual outputs are shaped by $\mathrm{pr}(\mathbf{p})$. For less than an infinite number of evaluations, the resulting probability is an estimate, denoted $\hat{\mathbb{P}}$, and given by the number of evaluations where one or more eigenvalues has a positive real part divided by J. Because $\mathbb{P}$ is a *binomial* variable (i.e. the outcome of each Monte Carlo evaluation takes one of two values: *stable* or *unstable*) exact confidence intervals for $\mathbb{P}$ are calculated using the binomial test (Conover, 1980). Confidence intervals also can be calculated for the difference $\Delta\mathbb{P}$ between $\mathbb{P}_1$ and $\mathbb{P}_2$ of two different control systems (Ray

* Received 28 May 1991; revised 12 February 1992; received in final form 16 June 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor P. Dorato.

† Department of Mechanical Engineering, Clemson University, Clemson, SC 29634, U.S.A.

‡ Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, U.S.A.

and Stengel, 1990):

$$\Pr[(L_1 - U_2) \le \Delta\mathbb{P} \le (U_1 - L_2)] \ge 1 - \alpha_1 - \alpha_2 + \alpha_1\alpha_2. \quad (1)$$

(L_1, U_1) and (L_2, U_2) are the binomial confidence intervals for $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively, and α_1, α_2 are the individual confidence coefficients. Confidence intervals for $\mathbb{P}$ are detailed in Stengel and Ray (1991).

The presentation of a probability wih exact confidence intervals is precise (nonconservative), defensible, and repeatable in the context of probability and statistics, even though it differs from accepted deterministic measures. Although the accuracy of $\mathbb{P}$ in describing the true robustness depends on the accuracy with which the parameters are described, known or estimated parameter uncertainties are required to determine whether a system is robust for any robustness measure (deterministic or stochastic). While deterministic measures such as a stability margin (Vicino *et al.*, 1990) can be computed without characterizing the uncertainty, the control system cannot be pronounced robust until the stability margin is evaluated with respect to the uncertainties expected in the system. Otherwise, the control system may be too robust, at the expense of performance, or nonrobust in the face of real-world uncertainties. By considering only allowed uncertainties, SRA eliminates this extra step. For deterministic measures, the problem is compounded by the fact that the metric itself can be conservative and/or difficult to compute.

Deterministic evaluation of the probability of instability. Continuously distributed parameters have a true underlying probability of instability that remains unknown for $J < \infty$, although given a sufficient number of evaluations, $\mathbb{P}$ can be bounded within a desired confidence interval (Stengel and Ray, 1991). Quantized distributions approximate continuous distributions, approaching them in the limit as the number of discrete parameter values goes to infinity. When a continuous distribution is approximated by a quantized distribution, it may be possible to perform fewer *deterministic* evaluations and obtain an equally good estimate $\hat{\mathbb{P}}$. For few parameters (r) and few quantization levels (w), deterministic evaluation may be a valid option. If the parameter distributions are in fact discrete, then evaluation of the w^r deterministic combinations would give $\mathbb{P}$, while Monte Carlo evaluation provides an estimate $\hat{\mathbb{P}}$ along with confidence intervals for $\mathbb{P}$. If r and/or w are large, then many deterministic evaluations are required, and Monte Carlo analysis may provide adequate bounds for fewer evaluations. The tradeoff between Monte Carlo analysis and deterministic evaluation depends on the number of parameters, their distributions, the number of quantization levels, and required confidence intervals. For example, a system with two binary parameters and $\mathbb{P} = 0.25$ requires 2^2 or four deterministic evaluations to obtain $\mathbb{P}$, but 6238 Monte Carlo evaluations are needed to compute 95% confidence intervals with a width of 10% of $\mathbb{P}$. A system with 20 binary parameters and $\mathbb{P} = 0.25$ requires 2^{20} or over 10^6 deterministic evaluations, yet its probability of instability can be bounded within 10% of $\mathbb{P}$ at a 95% confidence level with the same 6238 Monte Carlo evaluations. Evaluation of binary combinations is comparable to determining the stability of "corners" in parameter space. If the parameters distributions are not binary but the maximum real eigenvalue component is monotonic in the individual elements of $\mathbf{p}$, then the deterministic binary evaluations circumscribe results obtained for the actual continuous or quantized distributions with the same limits, and a conservative estimate $\hat{\mathbb{P}}$ is provided.

Extensions of stochastic stability robustness analysis

Stability robustness of systems with estimators. Stochastic stability robustness analysis is easily extended to systems incorporating dynamic state estimators. Using $\mathbf{F_A}$, $\mathbf{G_A}$, and $\mathbf{H_A}$ as the actual system matrices and $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ as the design system, the closed-loop system matrix for state ($\mathbf{x}$) and error dynamics ($\mathbf{x} - \hat{\mathbf{x}}$) is (Stengel, 1986)

$$\mathbf{F}_{cl} = \begin{bmatrix} \mathbf{F_A} - \mathbf{G_A}\mathbf{C} & -\mathbf{G_A}\mathbf{C} \\ (\mathbf{F} - \mathbf{F_A}) - (\mathbf{G} - \mathbf{G_A}) - \mathbf{K}(\mathbf{H} - \mathbf{H_A}) & \mathbf{F} - (\mathbf{G} - \mathbf{G_A})\mathbf{C} - \mathbf{KH} \end{bmatrix}, \quad (2)$$

where $\hat{\mathbf{x}}$ is the state estimate and $\mathbf{K}$ is the estimator gain matrix. The effect of parameter uncertainty on stability robustness is computed by Monte Carlo evaluation of the eigenvalues of equation (2), with $\mathbf{F}(\mathbf{p})$, $\mathbf{G}(\mathbf{p})$, and $\mathbf{H}(\mathbf{p})$ substituted for $\mathbf{F_A}$, $\mathbf{G_A}$, and $\mathbf{H_A}$. Closed-loop eigenvalue densities portrayed on the stochastic root locus show the possible interaction of dynamic and estimator state elements, and the possible robustness degradation due to the estimator. Well-known loss of LQ stability margins when a state estimator is added (Doyle, 1978) can be quantified by the probability of instability.

Performance robustness analysis. While stability is an important element of robustness, performance robustness analysis is vital to determining whether important design specifications are met. Stochastic stability robustness is described by a single parameter, the probability of instability. Adequate performance—initial condition response, response to commanded inputs, control authority, and rejection of disturbances is difficult to describe by a single scalar. However, elements of SRA (e.g. Monte Carlo evaluation and use of the binomial confidence intervals) apply, independent of the performance criteria chosen.

Numerous criteria stemming from classical control concepts exist as measures of adequate performance. Appealing to these, one can begin a smooth transition from stability robustness analysis to performance robustness analysis simply by analysing the *degree of stability* or *instability* rather than strict stability (Stengel and Ray, 1991). One method of doing this is to shift the vertical discriminant line from zero to Σ less than (or greater than) zero. Histograms and cumulative distributions for degree of stability are readily given by the Monte Carlo estimate of $\Pr(\Sigma)$, the probability that the maximum real eigenvalue component is less than Σ, where $-\infty < \Sigma < \infty$. Binomial confidence intervals are applicable to each point of the cumulative distribution, as there are just two values of interest, e.g. "satisfactory" or "unsatisfactory". The robustness measure resulting from the cumulative probability distribution is directly related to classical concepts of rates of decay (growth) of the closed-loop response, time-to-half and time-to-double. Rather than a vertical discriminant line, one can confine the closed-loop roots to *sectors in the complex plane* bounded by lines of constant damping and arcs of constant natural frequency. Systems with roots confined to sectors would be expected to display a certain transient response speed. Again, the probability of roots lying within a sector is a binomial variable, and binomial confidence interval calculations apply.

While the speed of the transient response depends on the closed-loop poles, its magnitude and overall shape depend on the coefficients of the characteristic exponential and sinusoidal terms, and time responses provide the most clear-cut means of evaluating performance. When time responses are computed, stochastic performance robustness can be portrayed as a distribution of possible trajectories around a nominal or desired trajectory. Envelopes can be defined around a nominal trajectory based on stated performance criteria, and the probability of exceeding the envelope becomes the scalar, binomial performance robustness metric (Ray and Stengel, 1990). While it is simple to conclude that a response violates such an envelope, individual responses within the envelope may not be acceptable. In such cases, the derivative of a response and envelopes around the derivative also can be used to evaluate performance. There is no unique set of criteria defining envelopes that bound an acceptable response; the envelopes may be defined by segments connecting points based on minimum/maximum dead time, rise time, time to peak

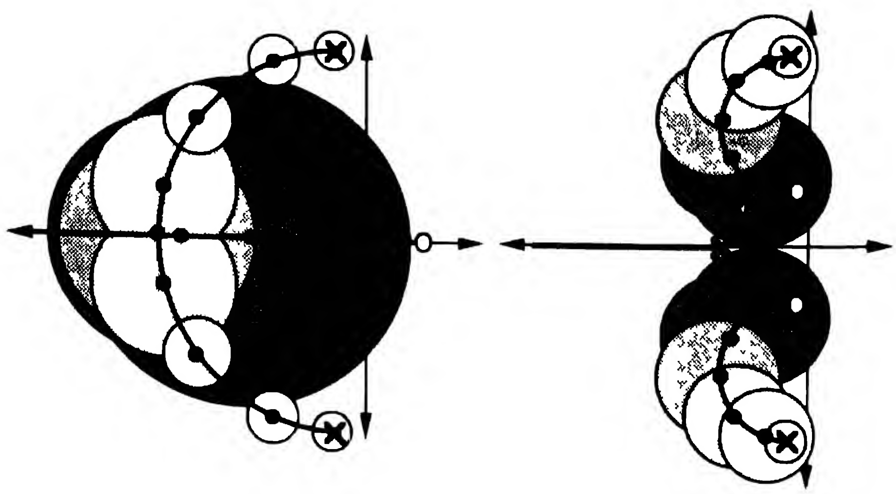


FIG 1 Illustration of design insight revealed by SRA Solid points indicate closed-loop eigenvalues enclosed by hypothetical ' uncertainty circles (a) Root locus where stability robustness decreases monotonically with increased gain (b) Root locus where stability robustness decreases, increases, then decreases with increased gain

overshoot, peak overshoot, settling time and steady state error Segmented envelopes can be smoothed or other scalars can be used to define points on the envelope However, once an envelope is defined time response distributions can be computed by Monte Carlo methods The closed-loop time response to a command input disturbance initial condition, or some combination is evaluated J times and for each evaluation the trajectory is a binomial variable, it either stays within the envelope or violates the envelope Although computing time responses are more computation intensive than probability of instability or degree of stability estimation such analysis is well within the capability of existing workstations

Stochastic robustness as a control design aid While general "rules of thumb' regarding the design of robust control systems are useful, SRA can identify non obvious robustness behavior in particular problems Figure 1 provides one example Consider Fig 1a which shows the root locus of a system that has a complex pair of poles and a right half plane zero Hypothetical bounded uncertainty circles are drawn around possible closed loop root locations the uncertainty circle at the pole represents possible open loop eigenvalue locations due to uncertainty in the dynamic matrix F As gain increases along the root locus the uncertainty is magnified and uncertainty in the control effect matrix contributes to the widening circles Stability robustness decreases and closed loop roots may be in the right-half plane at high enough gain This case illustrates one where the decrease in robustness is monotonic, as indicated by plotting the probability of instability vs the root-locus gain in Fig 2a Figure 1b postulates a system with a real pole and a complex pair of poles and zeros located in a 'pole over zero configuration The complex portion of the root locus starts near the $j\omega$ axis before ending at the zero in the left half plane Again, uncertainty circles enlarge as gain increases In this case, it is possible that eigenvalue distributions cross into the right-half plane are entirely in the left half plane as gain increases, and finally, cross back into the right-half plane at very high gain Here, stability robustness (as measured by the probability of instability) may have local or global minima as functions of gain (Fig 2b) For multivariable systems with many parameters, the intrinsic structure of the problem and the tradeoff between the spread in closed loop root unc rtainty vs the magnitude of the control gains may not be immediately evident Plots of stochastic robustness metrics vs scalar controller design parameters provide the necessary insight

Case study LQG/LTR system robustness analysis

It is well known that the stability margins guaranteed for an LQ system are not retained in LQG systems (Doyle, 1978) Nevertheless Loop Transfer Recovery (LTR) (Doyle *et al* 1979) has become an established method of recovering transfer properties of the LQ system (or the linear-optimal estimator, Kwakernaak (1976)) in minimum-phase LQG systems The condition for recovery of LQ transfer properties is derived from the fact that if $\mathbf{u}(t)$ had the same effect on both $\mathbf{x}(t)$ and $\hat{\mathbf{x}}(t)$, then the LQG system would have the same transfer function properties as the LQ system If the actual system matrices match those used to design the estimator, the transform relationships are (Doyle and Stein,

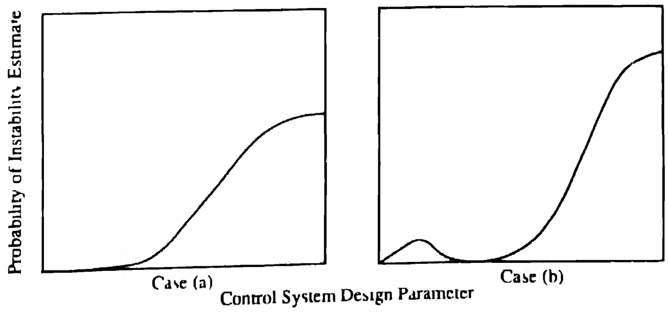


FIG 2 Probability of instability as a function of control system design parameter, for the two cases in Fig 1

1979)

$$\mathbf{x}(s) = (s\mathbf{I}_n - \mathbf{F})^{-1}\mathbf{G}\mathbf{u}(s), \tag{3}$$

$$\hat{\mathbf{x}}(s) = [(s\mathbf{I}_n - \mathbf{F}) + \mathbf{K}\mathbf{H}]^{-1}[\mathbf{K}\mathbf{H}(s\mathbf{I}_n - \mathbf{F})^{-1}\mathbf{G}\mathbf{u}(s) - \mathbf{G}\mathbf{u}_c(s)]. \tag{4}$$

When the estimator gain is chosen according to the recovery condition $\mathbf{K}/v = \mathbf{G}$, equation (4) becomes,

$$\hat{\mathbf{x}}(s) = \mathbf{A}^{-1}\mathbf{G}(\mathbf{H}\mathbf{A}^{-1}\mathbf{G})^{-1}\mathbf{H}\mathbf{A}^{-1}\mathbf{G}\mathbf{u}(s) = \mathbf{A}^{-1}\mathbf{G}\mathbf{u}(s) = \mathbf{x}(s), \tag{5}$$

where $\mathbf{A} = (s\mathbf{I} - \mathbf{F})$. Asymptotic recovery occurs as the positive, scalar design parameter v approaches ∞, and q estimator eigenvalues approach the q transmission zeros of $\mathbf{H}(s\mathbf{I}_n - \mathbf{F})^{-1}\mathbf{G}$. The procedure recovers the original loop only if the LQ system has minimum-phase transmission zeros. To meet the recovery condition, the nominal disturbance spectral density matrix, $\mathbf{W}_0$, is appended as

$$\mathbf{W} = \mathbf{W}_0 + v^2\mathbf{G}\mathbf{G}^T. \tag{6}$$

The term $v^2\mathbf{G}\mathbf{G}^T$ represents additional process noise that ruins the optimality of the estimator with respect to actual measurement noise and disturbances; hence, performance suffers as more process noise is added, but the good transfer function properties of LQ systems are recovered asymptotically. The tradeoff between estimator performance and system robustness is made by adjusting v.

Unstructured-singular-value analysis (USVA) typically is used to determine when LQ properties are recovered. Nevertheless, USVA does not indicate the effectiveness of LTR in systems with parameter uncertainties, as pointed out in Tahk and Speyer (1987) and Shaked and Soroka (1985). When the system description is uncertain, the actual system matrices do not match those used to design the estimator, and

$$\hat{\mathbf{x}}(s) = \mathbf{A}^{-1}\mathbf{G}[\mathbf{H}\mathbf{A}^{-1}\mathbf{G}]^{-1}[\mathbf{H}\mathbf{A}_\mathbf{A}^{-1}\mathbf{G}_\mathbf{A}]\mathbf{u}(s), \tag{7}$$

where $\mathbf{A}_\mathbf{A} = (s\mathbf{I} - \mathbf{F}_\mathbf{A})$ and $\mathbf{F}_\mathbf{A}$, $\mathbf{G}_\mathbf{A}$ represent the actual system matrices. Equation (7) shows that when parameter uncertainties are present, the original loop is not recovered, although partial recovery may improve robustness over that of the nominal LQG system. q estimator poles approach the transmission zeros as $v \to \infty$, and finite minimum phase transmission zeros may move to the right-half plane due to parameter uncertainty in $\mathbf{H}\mathbf{A}_\mathbf{A}^{-1}\mathbf{G}_\mathbf{A}$. Estimator poles also are influenced by parameter uncertainty and can vary around the transmission zero: hence, increasing v indefinitely can *decrease* robustness in systems with one or more uncertain parameters. In such systems, SRA determines the value of v required to recover sufficient robustness while maintaining adequate performance, as demonstrated by the example that follows. This characteristic of SRA is not limited to analysis of LQG/LTR system but is useful with any control design approach in which one or more design parameters is arbitrary.

Single-link robot arm. A flexible one-link robot typically is used to study problems associated with controlling a compliant system when the sensor and actuator are not collocated (e.g. Cannon and Schmitz, 1984). In such systems, robustness concerns can be severe. The linear model of the single-link robot arm retains the first three flexible modes, and the tip of the link is controled by applying a control torque to the hub, or base of the link. Because the model is representative of a general flexible structure, physical parameters are easily identifiable, and robustness is a concern, it is a good candidate for SRA.

The dynamic, control effect, and output matrices are given by

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\omega_1^2 & -2\zeta_1\omega_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\omega_2^2 & -2\zeta_2\omega_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\omega_3^2 & -2\zeta_3\omega_3 \end{bmatrix} \tag{8}$$

$$\mathbf{G} = \frac{1}{I_T}[0 \quad 1 \quad 0 \quad \phi_1'(0) \quad 0 \quad \phi_2'(0) \quad 0 \quad \phi_3'(0)]^T, \tag{9}$$

$$\mathbf{H} = \begin{bmatrix} L & 0 & \phi_1(L) & 0 & \phi_2(L) & 0 & \phi_3(L) & 0 \\ 0 & 1 & 0 & \phi_1'(0) & 0 & \phi_2'(0) & 0 & \phi_3'(0) \end{bmatrix}, \tag{10}$$

where x is the length along the arm, $\phi_i(x)$, are the normal modes, $\phi_i' = \frac{d\phi_i}{dx}$, L is the length of the arm, and I_T is the total inertia of the arm. The measurements taken through $\mathbf{H}$ are the tip displacement and hub-rate, respectively. The flexibility of the open-loop system is apparent in the open-loop eigenvalues, which are 0, 0, $-0.177 \pm 11.81j$, $-0.432 \pm 21.61j$, and $-0.968 \pm 48.37j$. The transfer function between tip displacement and hub torque is non-minimum phase, with zeros 12.4, -12.0, $21.6 \pm 24.2j$, $-22.5 \pm 24.2j$; hence, a non-minimum phase response can be expected for tip displacement. The system has a readily identifiable 14-element parameter vector:

$$\mathbf{p} = [\zeta_1 \quad \omega_1 \quad \zeta_2 \quad \omega_2 \quad \zeta_3 \quad \omega_3 \quad \phi_1'(0) \quad \phi_2'(0) \quad \phi_3'(0) \quad L \quad \phi_1(L) \quad \phi_2(L) \quad \phi_3(L) \quad I_T]. \tag{11}$$

Details concerning the modeling and parameter identification are given in Cannon and Schmitz (1984). The linear-quadratic regulator designed in Cannon and Schmitz (1984) is used for demonstration of SRA. The performance index weights tip position and tip rate, and the LQR state-weighting, control-weighting, and control gain matrices are

$$\mathbf{Q} = 0.01\mathbf{F}^T\mathbf{H}^T\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{H}\mathbf{F} + \mathbf{H}^T\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{H}, \quad \mathbf{R} = 0.001, \tag{12)-(13}$$

$$\mathbf{C} = [35.42 \quad 13.38 \quad 41.24 \quad 2.65 \quad 59.32 \quad -0.67 \quad 135.46 \quad 1.58]. \tag{14}$$

The nominal closed-loop eigenvalues are $-5.41 \pm 48.8j$, $-6.47 \pm 23.8j$, $-6.1 \pm 2.66j$, $-7.7 \pm 11.42j$.

A uniform probability density function models the parameter uncertainties, with variations between $\pm 2\%$ of the nominal values for L and I_T and $\pm 25\%$ for the remaining parameters. The 50,000-evaluation stochastic root-locus for the full-state feedback system is given in Fig. 3. The nominal eigenvalues are marked, and the distribution is indicated by the height above the complex plane in units of roots/length along the real axis and roots/area in the complex plane. The "bin" size in Fig. 3 is 0.9 along the axis and 0.9×0.9 off the axis. For 50,000 evaluations, $\hat{\mathbb{P}}$, is zero, with 95% confidence intervals $(L, U) = (0, 7.4 \times 10^{-5})$. Each of the four complex eigenvalue pairs appears in Fig. 3 as a "peak", with a surrounding distribution due to parametric uncertainty. The peaks can be well-defined (as in the lowest frequency complex pair) or broad (as in the highest frequency pair) and the nominal eigenvalues are not necessarily at the distributions' peak. Parameter uncertainty causes complex pairs to coalesce into real roots resulting in a distribution along the real axis. The closed-loop eigenvalues tend to spread into the left-half plane, while definite boundaries are delineated on the right. For binary parameter variations of the same magnitude as the maximum uniform variations, 2^{14} or 16,384 deterministic evaluations also give $\mathbb{P} = 0$. These results indicate good stability robustness in the face of reasonably large uncertainties.

Moving to performance robustness analysis, Fig. 3 shows sector bounds defined by $4 \le \omega_n \le 65$ and $\zeta > 0.1$. For 50,000 evaluations, the probability that closed-loop eigenvalues lie outside of these bounds is 0.0147, with 95% confidence intervals $(L, U) = (0.0136, 0.0158)$. While the shape of the time response depends on closed-loop zeros, a minimum response speed can be guaranteed by requiring that all closed-loop eigenvalues lie within the specified sector.

Figure 4(a) presents segmented step-response envelopes and 500 Monte Carlo evaluations of the response of the tip to a 4.8 cm position command input. The control history corresponding to the mean response is given in Fig. 4b. The

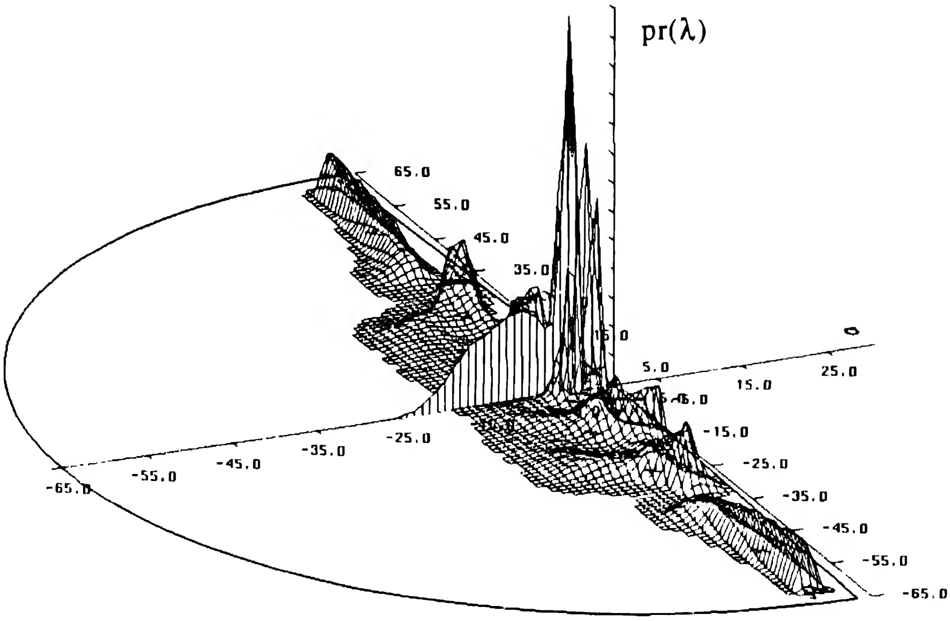


FIG. 3. Stochastic root-locus for the single-link robot with uniform parameters, 50,000 evaluations. Nominal eigenvalues are marked by 'x'.

initial response is in the wrong direction, since the transfer function is non-minimum phase; the envelopes in Fig. 4 indicate the maximum acceptable non-minimum phase response. For 500 responses, the probability of violating the time response envelope is 0.184 with 95% confidence intervals $(L, U) = (0.151, 0.221)$. Individual responses characteristic of those evaluated by Monte Carlo analysis are given in Fig. 5. While responses fill out the envelope, some of the individual responses within the envelope may not be acceptable in the face of real-world criteria governing rate of change of the response (Fig. 5c). This is a case where checking envelopes around the derivative of the response may be necessary. Similar analyses can be performed on control trajectories to make sure bandwidth and control effort limitations are not violated.

It is instructive from a design standpoint to plot robustness measures vs design parameters used to calculate feedback gains. Since there is a single control in this example, the scalar control weighting matrix R can be used as the design parameter. Two stochastic performance robustness measures are plotted vs R in Fig. 6—the probability of violating the time-response envelope and the probability of degree of instability. As control gains increase, the closed-loop roots are pushed farther into the left-half plane, but they also tend to migrate farther from their nominal values. At some value of control gain, there is a tradeoff between how far roots migrate and their location in

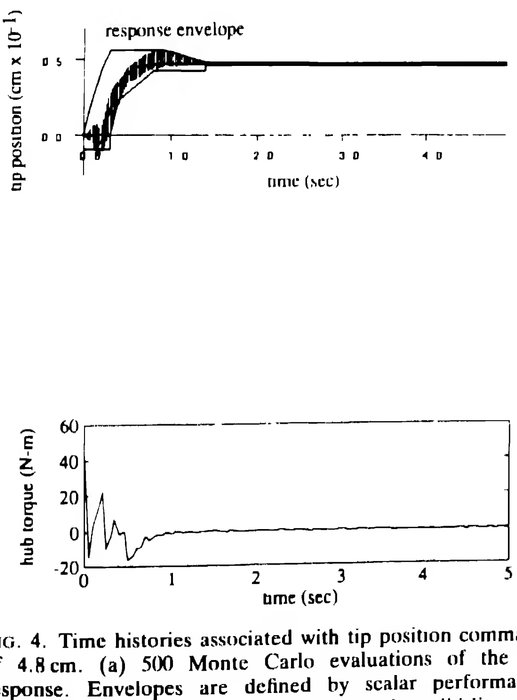


FIG. 4. Time histories associated with tip position command of 4.8 cm. (a) 500 Monte Carlo evaluations of the tip response. Envelopes are defined by scalar performance criteria. Nominal response is indicated by the solid line. (b) Nominal control input.

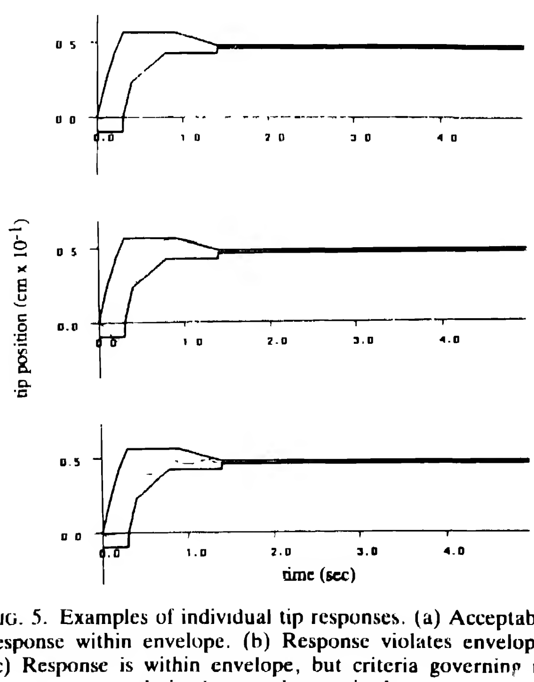


FIG. 5. Examples of individual tip responses. (a) Acceptable response within envelope. (b) Response violates envelope. (c) Response is within envelope, but criteria governing its derivative may be required.

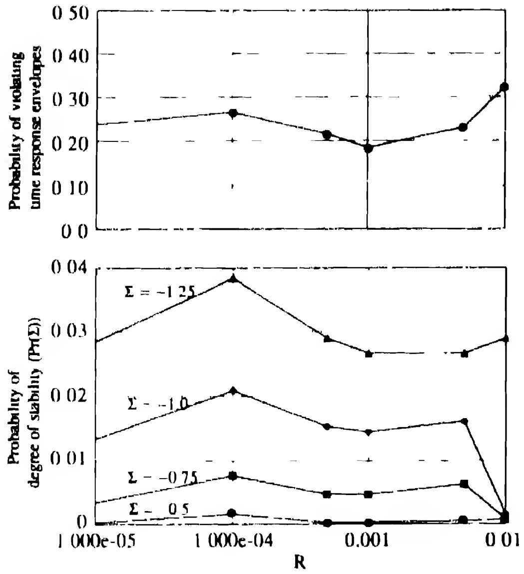


FIG. 6. Stochastic performance robustness metrics vs control weighting matrix **R**. (a) Probability of violating time response envelopes. (b) Probability of degree of instability, for values Σ along the real axis.

the left-half plane; thus a local minimum is apparent around $R = 0.001$ in the probability-of-degree-of-instability curves. While degree of instability improves for very small R, the control gains become unrealistically large. For larger R (smaller control gains), the nominal closed-loop roots have real parts in the range of the values of Σ used; hence, the probability of degree of instability increases rapidly beyond $R = 0.01$. The probability of violating time-response envelopes and the probability of degree of instability show similar trends—a broad minimum in the region $R = 0.001$—as functions of R.

Figure 7 shows the stochastic root-locus for the LQG system with estimator gains based on disturbance effect matrix $\mathbf{L} = \mathbf{G}$, and disturbance and noise covariance matrices $W = 1\ (\mathrm{N \cdot m})^2$, $\mathbf{N} = \mathrm{diag}\,[0.005\,\mathrm{cm}^2\ \ 10\ (\mathrm{rad\,sec}^{-1})^2]$. The stochastic root-locus of the LQG system changes in overall character from that of the LQ system. Peaks are sharper, and the real-axis distribution is less pronounced. In particular, note the eigenvalues associated with the largest peaks. In Fig. 3 a broad distribution is associated with these eigenvalues, yet in Fig. 7, this pair of eigenvalues shows little variation from its nominal location! While the extent of the distribution into the left-half plane is about the same as in Fig. 3, LQG system eigenvalues do migrate into the right-half plane The probability-of-instability estimate and 95% confidence intervals for 50,000 evaluations are $\hat{P} = 0.0771$, and $(L, U) = (0.0748, 0.0795)$, representing a significant loss in the stability robustness characteristic of the LQ system. Figure 8 illustrates the effect of Loop Transfer Recovery for this example. Figure 8a shows that there is a value of ν $(\nu = 2)$ that minimizes the probability of instability. The fact that such a minimum exists and the value of the design parameter that minimizes the probability of instability are not apparent by simply examining the nominal eigenvalues of the LQG/LTR system. The finite transmission zeros of the system are $-1.75 \pm 46.8j$, $-2.9 \pm 18.2j$, $-5.8 \pm 6.6j$, and -6.5. As ν increases, seven eigenvalues approach the system transmission zeros; the minimum can be attributed to the tradeoff between uncertainty in eigenvalue location as gains increase and the nearness of transmission zeros to the imaginary axis. While a minimum probability of instability is not guaranteed, the results presented in Fig. 8a offer design insight and show robustness characteristics that may not be revealed by deterministic robustness measures, particularly if the deterministic analysis is conservative. The tradeoff between estimator performance and system robustness is shown by comparing Figs 8a and b. Figure 8b

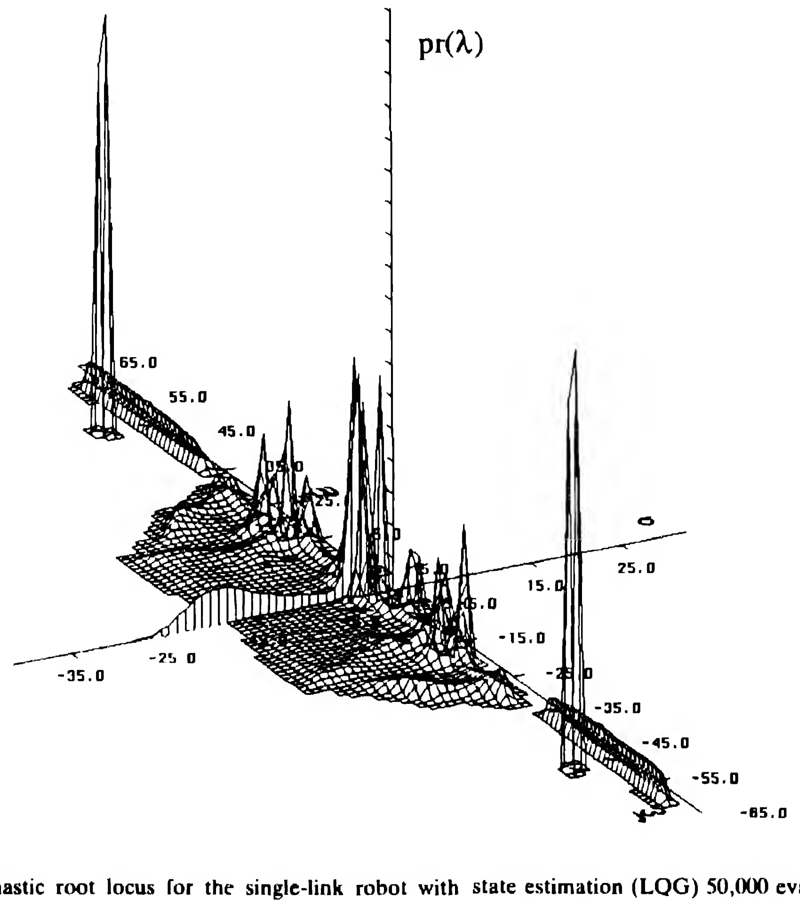


FIG. 7. Stochastic root locus for the single-link robot with state estimation (LQG) 50,000 evaluations

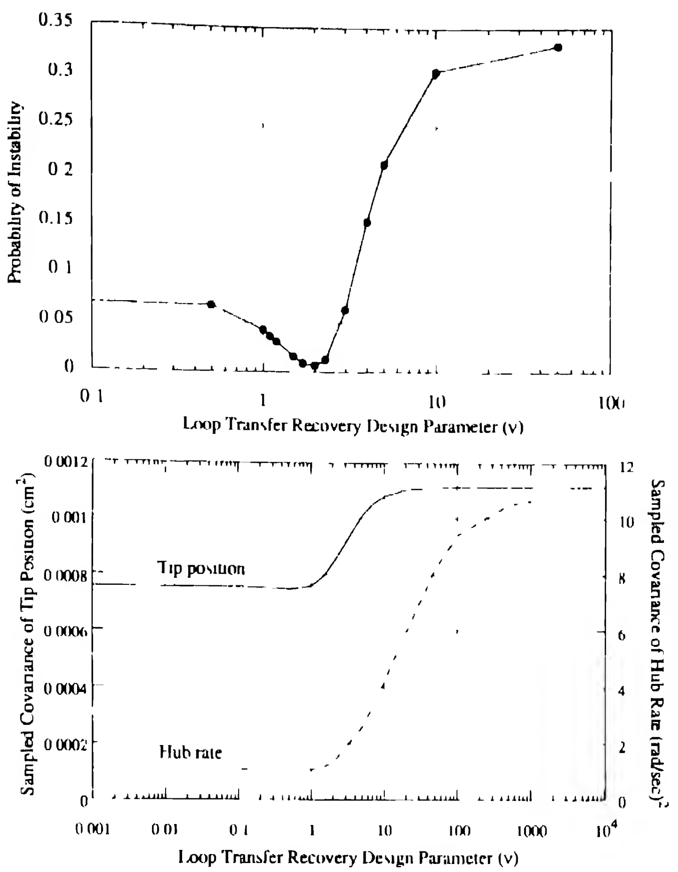


FIG. 8. Evaluation of Loop Transfer Recovery (a) Probability of Instability vs LQG/LTR design parameter v. (b) Sampled estimate of tip position and hub rate covariances vs LQG/LTR design parameter v.

indicates estimator performance by sampled estimates of the covariance of the output, $\mathbf{P} = \mathbf{H}E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^2]\mathbf{H}^T$, where $E[\cdot]$ is the expectation operator. The output covariance (based on simulation of the LQG system) shows that performance degradation over that of the nominal LQG system is small at the minimizing value of v

Conclusion

Stochastic Robustness Analysis offers a rigorous yet straightforward alternative to current metrics for control system robustness that is simple to compute and is unfettered by normally difficult problem statements, such as non-Gaussian statistics, arbitrary functions of uncertain parameters appearing as matrix elements, and structured uncertainty. Principles behind stochastic robustness can be applied to scalar performance metrics and/or time responses, making it a good candidate for overall robustness analysis. Stability and performance measures resulting from the analysis can provide details relating intrinsic robustness characteristics and control system design parameters. The analysis makes good use of the computing power of modern workstations. The example demonstrates stochastic robustness analysis applied to LQG/LTR. The analysis determines the effectiveness of Loop Transfer Recovery on uncertain systems, and the Loop Transfer Recovery design parameter that gives adequate stability robustness with minimal performance degradation is readily identified.

Acknowledgement—This research has been sponsored by the FAA and the NASA Langley Research Center under Grant No. NGL 31-001-252 and by the Army Research Office under Grant No. DAAL03-89-K-0092.

References

Cannon, R. J and E. Schmitz (1984). Initial experiments on the end-point control of a flexible one-link robot. *Int. J. of Robotics Res.*, **3**, 62–75.

Conover, W J. (1980). *Practical Non-parametric Statistics.* John Wiley, New York.

Doyle, J C. (1978). Guaranteed margins for LQG regulators *IEEE Trans. on Aut. Control*, **AC-23**, 756–757.

Doyle, J. C. and G Stein (1979). Robustness with observers. *IEEE Trans on Aut. Control*, **AC-24**, 607–611.

Doyle, J C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proceedings*, **129**, 242–250.

Kwakernaak, H. (1976). Optimal low-sensitivity linear feedback systems. *Automatica*, **5**, 378–382.

Lehtomaki, N. A., N. R. Sandell Jr. and M. Athans (1981). Robustness results in linear quadratic Gaussian-based multivariable control designs. *IEEE Trans. on Aut. Control*, **AC-26**, 75–93.

Ray, L. R. and R. F Stengel (1990). Stochastic stability and performance robustness of aircraft control systems. *AIAA Guidance, Navigation, and Control Conf.*, Portland, OR, pp. 863–873.

Ray, L. R. and R. F. Stengel (1991). Application of stochastic robustness to aircraft control systems. *AIAA J. of Guidance, Control, and Dynamics*, **14**, 1251–1259.

Siljak, D. D. (1989). Parameter space methods for robust control design: a guided tour. *IEEE Trans. on Aut. Control*, **AC-34**, 674–688.

Shaked, U and E. Soroka (1985). On the stability of the continuous-time LQG optimal control. *IEEE Trans. on Aut. Control*, **AC-30**, 1039–1043.

Stengel, R. F (1980). Some effects of parameter variations

on the lateral-directional stability of aircraft. *AIAA J. of Guidance and Control*, **3**, 124–131.

Stengel, R. F. (1986). *Stochastic Optimal Control: Theory and Application*. John Wiley, New York.

Stengel, R. F. and L. R. Ray, (1991). Stochastic robustness of linear time-invariant systems. *IEEE Trans. on Aut. Control*, **36**, 82–87.

Tahk, M. and J. L. Speyer (1987). Modeling of parameter variations and asymptotic LQG synthesis. *IEEE Trans. on Aut. Control*, **AC-32**, 793–801.

Vicino, A., A. Tesi and M. Milanese (1990). Computation of nonconservative stability perturbation bounds for systems with nonlinearly correlated uncertainties. *IEEE Trans. on Aut. Control*, **35**, 835–841.

Automatica, Vol. 29, No. 1, pp. 237–240, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Robust Root Clustering for Linear Uncertain Systems Using Generalized Lyapunov Theory*

R. K. YEDAVALLI†

Key Words—Robust pole-placement; root clustering; robust control, linear state space systems; unstructured uncertainty; structured uncertainty; Lyapunov theory.

Abstract—In this paper, the problem of matrix root clustering in sub-regions of complex plane for linear state space models with real parameter uncertainty is considered. The nominal matrix root clustering theory of Gutman and Jury (1981, *IEEE Trans. Aut. Control*, **AC-26**, 403) using Generalized Lyapunov Equation is extended to the perturbed matrix case and bounds are derived on the perturbation to maintain root clustering inside a given region. The theory allows us to get an explicit relationship between the parameters of the root clustering region and the uncertainty range of the parameter space. The current literature available on perturbation bounds for robust stability becomes a special case of this unified theory.

1. *Introduction*

THE PROBLEM OF ANALYZING and designing controllers for linear systems subject to real parameter uncertainty has been an extremely active topic of research in recent years. For example, see Dorato and Yedavalli (1990) and Siljak (1989) for a summary of recent developments in this area. In particular, there is considerable literature available on the analysis of robustness of linear state space systems with real parameter perturbations where the uncertainty can be either norm bounded (unstructured) or in terms of bounds on the intervals of the parameters (structured). However most of the analysis is essentially devoted to the robust stability problem wherein the stability region is the entire open left half of the complex plane for continuous time systems and the unit circle with center at the origin for discrete time systems. The more general problem of robust D-stability where a 'D-region' is any given subregion in the complex plane has received much less attention. Since 'pole-placement' technique is an effective way of shaping the dynamical response, both for continuous as well as discrete time systems, robust D-stability problem is essentially a performance robustness problem in which the stability robustness problem becomes a special case. Henceforth the phrases *robust* D-*stability, robust root clustering, robust eigenvalue placement* will be used interchangeably.

Most of the literature on robust D-stability is confined to family of polynomials (Barmish, 1989; Fu and Barmish, 1989; Soh, 1989; Kokame and Mori, 1991; Ackermann *et al.*, 1991; Vicino, 1989; Zeheb, 1989). The very few methods reported for *matrix* root clustering confine themselves to some very specific D-regions (Juang *et al.*, 1989; Juang, 1991; Sobel and Yu, 1989; Keel *et al.*, 1991; Tesi and Vicino, 1990). In majority of these papers, the relationship between perturbation range and the eigenvalue migration range is not explicit and is not tractable. In this paper an elegant, unified theory for robust eigenvalue placement is presented for a class of D-regions defined by algebraic inequalities by extending the nominal Matrix Root Clustering theory of Gutman and Jury (1981) to linear uncertain systems (valid for both continuous and discrete time systems; both for structured and unstructured uncertainties; as well as analysis and design). Incidentally, this type of extension was considered in a series of papers by Abdul-Wahab (1990, 1991) with continuous time systems in mind. But as pointed out by Yedavalli (1992) the results were erroneous. We present here explicit conditions for matrix root clustering for different D-regions (which in turn have direct effect on the time response of the system) in terms of bounds on the parameter perturbations and establish the relationship between eigenvalue migration range and parameter range which are valid for both continuous time as well as discrete time systems. The bounds obtained do not need any frequency sweeping or parameter gridding.

The paper is organized as follows. In Section 2, we briefly review the nominal matrix root clustering theory of Gutman and Jury (1981) using Generalized Lyapunov theory. Section 3 develops the relationship between parameter perturbation range and parameters of the root clustering region for different regions. In Section 4 we illustrate the theory with an example and finally Section 5 offers some concluding remarks.

2. *Root clustering theory for a nominal matrix*

In what follows, we essentially use the same notation followed by Gutman and Jury in their paper. In their paper they consider a two variable transformation region Ω for matrix root clustering. For simplicity in exposition we restrict our attention to only real matrices and review the material in Gutman and Jury (1981) related to only real matrices.

Let $A \in R^{n\times n}$, λ be the eigenvalue of A, $\bar{\lambda}$, the complex conjugate of λ, $x = \mathrm{Re}[\lambda]$ and $y = \mathrm{Im}[\lambda]$. For $A \in R^{n\times n}$, we consider a region symmetrical about the real axis, described by the algebraic inequalities

$$\Omega_v = \left\{(x, y) : \sum_{f,g} \gamma_{fg} x^f y^g < 0\right\},$$

$$\dot{\Omega}_v = \left\{(x, y) : \sum_{f,g} \gamma_{fg} x^f y^g \leq 0\right\},$$

where $v = f + g$, $g = 2h$, f, g and h are nonnegative integers and v is the region's degree and γ_{fg} is a real coefficient. Note that $\bar{\Omega}$ simply includes the boundary of Ω_v. The following facts are reproduced from Gutman and Jury (1981).

$$x = \tfrac{1}{2}(\lambda + \bar{\lambda}),$$

$$y = -\frac{i}{2}(\lambda - \bar{\lambda}),$$

$$\mu(\lambda, \bar{\lambda}) = \sum_{f,g} \gamma_{fg}(-1)^h(\tfrac{1}{2})^{f+2h} \times (\lambda + \bar{\lambda})^f(\lambda - \bar{\lambda})^{2h} = \sum_{f,g} \gamma_{fg} x^f y^{2h}.$$

* Received 2 October 1991; received in final form 29 May 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor J. E. Ackermann.

† Department of Aeronautical and Astronautical Engineering, The Ohio State University, Columbus, OH 43210-1276, U.S.A.

In addition, let α, $\beta \in C$ and define

$$\mu(\alpha, \beta) = \sum_{f,g} \gamma_{fg}(-1)^h(\tfrac{1}{2})^{f+2h}(\alpha+\beta)^f(\alpha-\beta)^{2h}. \quad (1)$$

For simplicity in exposition, we limit our attention to regions Ω_1 and Ω_2 and specialize the above notation to these two regions. Incidentally, these two regions cover quite a large class of regions in the complex plane. The following are examples of a class of regions:

Regions of Degree 1.

$$\Omega_1 : \{(x, y) : \gamma_{00} + \gamma_{10}x < 0\}.$$

These regions include open left half plane and regions with prescribed degree of stability.

Regions of Degree 2.

$$\Omega_2 : \{(x, y) : \gamma_{00} + \gamma_{10}x + \gamma_{02}y^2 + \gamma_{20}x^2 < 0\}.$$

This represents a conic section (either ellipse, parabola or hyperbola, depending on the nature of the coefficients γ_{ij}s).

We now recall a fundamental theorem on root clustering of a nominal matrix in terms of Generalized Lyapunov Equation (G.L.E.) from Gutman and Jury (1981). Consider the Genralized Lyapunov Equation

$$\sum_{p,q} c_{pq}A^pPA^{Tq} = -Q,$$

where A^T is the transpose of A and c_{pq} is the coefficient of $\alpha^p\beta^q$ in the polynomial $\mu(\alpha, \beta)$ given by (1).†

Note that for the regions under consideration coefficients c_{pq} are real. Before proceeding to state an important theorem found in Gutman and Jury (1981), in what follows, we summarize the expressions for c_{pq} and the expressions for the Generalized Lyapunov Equation for four regions, namely, LHP, α-shift, ellipse and circle.

- Open left half plane.

$$\Omega_1 : \{x < 0\}(\gamma_{00} = 0, \gamma_{10} = 1), \quad c_{00} = 0, \quad c_{10} = c_{01} = \tfrac{1}{2}, \quad (2)$$

$$\text{G.L.E.:} \quad (PA^T + AP) = -2Q. \quad (3)$$

- α degree of stability.

$$\Omega_1 : \{\alpha + x < 0, \alpha > 0\}, \quad (\gamma_{00} = \alpha, \gamma_{10} = 1),$$
$$c_{00} = \alpha, \quad c_{10} = c_{01} = \tfrac{1}{2}, \quad (4)$$

$$\text{G.L.E.:} \quad 2\alpha P + PA^T + AP = -2Q. \quad (5)$$

- Ellipse.

$$\Omega_2 : \{\gamma_{00} + \gamma_{02}y^2 + \gamma_{10}x + \gamma_{20}x^2 < 0\}(\gamma_{20} > 0, \gamma_{02} > 0),$$
$$c_{00} = \gamma_{00}, \quad c_{10} = c_{01} = \tfrac{1}{2}\gamma_{10}, \quad (6)$$
$$c_{11} = \tfrac{1}{2}(\gamma_{20} + \gamma_{20}), \quad c_{02} = c_{20} = \tfrac{1}{4}(\gamma_{20} - \gamma_{02}),$$

$$\text{G.L.E.:} \quad c_{00}P + c_{01}(PA^T + AP) + c_{11}APA^T + c_{02}(PA^{T2} + A^2P) = -Q. \quad (7)$$

- Circle.

$$\Omega_2 : \{\gamma_{00} + \gamma_{10}x + x^2 + y^2 < 0\},$$
$$c_{00} = \gamma_{00}, \quad \tfrac{1}{2}\gamma_{10}, \quad c_{11} = 1, \quad c_{02} = c_{20} = 0, \quad (8)$$

$$\text{G.L.E.:} \quad c_{00}P + c_{01}(PA^T + AP) + APA^T = -Q. \quad (9)$$

We now state the theorem on root clustering using Generalized Lyapunov Equation given in Gutman and Jury (1981).

***Theorem* 1 (see Gutman and Jury, 1981).** Let $A \in R^{n\times n}$ and consider Ω_2 given by (6) with $\gamma_{02} + \gamma_{20} \geq 0$. For the eigenvalues of A to lie in Ω_2 it is necessary and sufficient that given any positive definite matrix Q, there exists a unique positive definite symmetric matrix P satisfying (7) (with (6)).

† Note that there is no loss of generality if one considers the equation $\sum c_{pq}A^{Tp}PA^q = -Q$.

3. *Bounds for robust root clustering*

In this section, we extend the concepts of root clustering given in Gutman and Jury (1981) to perturbed matrices and derive bounds on the perturbation to maintain root clustering in a given region (robust root clustering). Towards this direction, we consider systems with both unstructured perturbation as well as structured perturbation.

Bounds for unstructured perturbation. Consider the following linear state space model

$$\dot{x} = Ax = (A_0 + E)x, \quad x(0) = x_0,$$

where A_0 is an $n \times n$ matrix with a given root clustering region and E is an unstructured perturbation on A_0.

The aim is to derive bounds on the norm of the perturbation of matrix, i.e. on $\|E\|$ such that $A_0 + E$ has roots maintained inside the root clustering region of A_0. Note that in a design situation, the matrix A_0 may represent a nominal closed loop system matrix with gain matrix elements as design parameters (for either continuous time or discrete time systems).

First consider the generalized Lyapunov equation of (5) corresponding to region of degree 1. Assuming that the eigenvalues of the nominal system matrix A_0 are located inside the given region Ω_1 (LHP or α-shifted LHP), we now want to derive bounds on the perturbation matrix E such that the roots of the perturbed system matrix $A_0 + E$ also lie inside the region Ω_1.

Theorem 2. The perturbed system matrix $A_0 + E$ has eigenvalues inside the given region Ω_1 of (4) if

$$\sigma_{\max}(E) < \frac{\sigma_{\min}(Q)}{\sigma_{\max}(P)} = \mu_{1_r},$$

where P satisfies

$$2\alpha P + PA_0 + A_0^TP = -2Q.$$

Proof. Similar to the proof given in Patel and Toda (1980)

Remark 1. Note that this bound μ_{1_r} specializes to the standard left half plane (asymptotic stability for continuous time systems) bound derived in Patel and Toda (1980) where $\alpha = 0$. Here μ_{1_r} denotes the perturbation bound for root clustering for region of degree 1.

Now consider the generalized Lyapunov equation of (7) corresponding to region of degree 2. Assuming that the eigenvalues of the nominal system matrix A_0 are located inside the given region Ω_2, we now want to derive bounds on the perturbation matrix E such that the roots of the perturbed matrix $A_0 + E$ also lie inside the region Ω_2.

Theorem 3. The perturbed system matrix $A_0 + E$ has eigenvalues inside the given region Ω_2 of (6) if

$$\sigma_{\max}(E) < \left[\left(b + \frac{c_{01m}}{a}\right)^2 + \frac{c}{a}\right]^{1/2} - \left(b + \frac{c_{01m}}{a}\right) = \mu_{2_r}$$

where

$$a - 2c_{02m} + c_{11m},$$
$$b = \sigma_{\max}(A_0),$$
$$c = \sigma_{\min}(Q)/\sigma_{\max}(P),$$

and P satisfies the generalized Lyapunov equation

$$c_{00}P + c_{01}(PA_0^T + A_0P) + c_{11}A_0PA_0^T + c_{02}(PA_0^{T2} + A_0^2P) = -Q,$$

and μ_{2_r} denotes the perturbation bound for root clustering for the region of degree 2 and $(\cdot)_m$ denotes the absolute value of $(\cdot)$.

Proof. See Appendix A.

For the special case of a circle in the left half plane with center at β and radius r_c, the generalized Lyapunov equation is given by the following parameters

$$c_{00} = \beta^2 - r_c^2, \quad c_{01} = c_{10} = -\beta, \quad c_{11} = 1, \quad c_{02} = c_{20} = 0.$$

Thus we have the G.L.E. as

$$-\beta(A_0P + PA_0^T) + A_0PA_0^T + (\beta^2 - r_c^2)P = -Q.$$

The above equation can be written as

$$\frac{(A_0 - \beta I_n)}{r_c} P \frac{(A_0 - \beta I_n)^T}{r_c} - P = -\frac{Q}{r_c^2},$$

which is in the form of a Discrete Lyapunov Equation with the nominal matrix $(A_0 - \beta I_n)/r_c$. For this case the bound μ_{2_n} is given by

$$\sigma_{\max}(E) < \mu_{2_n} = \left[(\sigma_{\max}(A_0) - \beta)^2 + \frac{\sigma_{\min}(Q)}{\sigma_{\max}(P)}\right]^{1/2} - (\sigma_{\max}(A_0) - \beta). \quad (10)$$

Remark 2. It may be noted that the bound μ_{2_n} specializes to the discrete system bounds of Kolla *et al.* (1989) with $\beta = 0$, $r_c = 1$.

Bounds for structured perturbation. For this case, we consider the linear state space system with structured perturbation as follows:

$$\dot{x} = A(q)x, \quad x(0) = x_0,$$

where

$$A(q) = A_0 + E(q) = A_0 + \sum^r q_i E_i, \quad (11)$$

with $A_0 \in R^{n \times n}$ being the "nominal" matrix obtained at the nominal value of the uncertain parameter vector q (q_i, $i = 1, 2, \ldots, r$), i.e. $q^0 = 0$ and E_i are given constant matrices. This type of representation produces a "polytope" in the matrix space. A special case of interest is the so-called "interval matrix" family in which E_i are such that they contain a single nonzero element, at a different location in the matrix for each different i.

We now define a set of matrices with the following notation. Let $[\ ]_m$ denote the matrix with all its elements taking on absolute values of the elements of the matrix $[.]$. Also let $[.]_s$ denote the symmetric part of the matrix, i.e $([.] + [.]^T)/2$.

Again consider the generalized Lyapunov equation of (5) corresponding to region of degree 1. Assuming that the eigenvalues of the nominal system matrix A_0 are located inside the given region Ω_1 (LHP or α-shifted LHP), we now want to derive bounds on the perturbation matrix $E(q)$ such that the roots of the perturbed system matrix $A_0 + E(q)$ also lie inside the region Ω_1.

Theorem 4. The perturbed system matrix $A_0 + E(q)$ has eigenvalues inside the given region Ω_1 of (4) if

$$|q_j| < \frac{\sigma_{\min}(Q)}{\sigma_{\max}(\sum (P_i)_m)} = \mu_{1_{rs}},$$

where $P_i = (PE_i)_s$ and P satisfies

$$2\alpha P + PA_0 + A_0^T P = -2Q.$$

Proof. Similar to the proof of Theorem 2.

Remark 3. Note that this bound $\mu_{1_{rs}}$ specializes to the standard left half plane (asymptotic stability for continuous time systems) bound derived in Keel *et al.* (1988) and Zhou and Khargonekar (1987) where $\alpha = 0$. Here $\mu_{1_{rs}}$ denotes the perturbation bound for root clustering for region of degree 1 for structured uncertainty.

Now consider the generalized Lyapunov equation of (7) corresponding to region of degree 2. Assuming that the eigenvalues of the nominal system matrix A_0 are located inside the given region Ω_2, we now want to drive bounds on the perturbation parameters q_i such that the roots of the perturbed matrix $A_0 + E(q)$ also lie inside the region Ω_2.

Let

$$P_{ep} = \begin{bmatrix} (E_1E_1P)_s & (E_1E_2P)_s & \cdots & (E_1E_rP)_s \\ (E_1E_2P)_s & (E_2E_2P)_s & \cdots & (E_2E_rP)_s \\ \vdots & & & \vdots \\ (E_1E_rP)_s & (E_2E_rP)_s & \cdots & (E_rE_rP)_s \end{bmatrix},$$

$$P_{ee} = \begin{bmatrix} (E_1PE_1^T)_s & (E_1PE_2^T)_s & \cdots & (E_1PE_r^T)_s \\ (E_1PE_2^T)_s & (E_2PE_2^T)_s & \cdots & (E_2PE_r^T)_s \\ \vdots & & & \vdots \\ (E_1PE_r^T)_s & (E_2PE_r^T)_s & \cdots & (E_rPE_r^T)_s \end{bmatrix}$$

and

$$P_{aei} = (E_iPA_0^T)_s,$$
$$A_{oip} = (A_0E_iP)_s,$$
$$E_{iap} = (E_iA_0P)_s,$$
$$P_{ei} = (E_iP)_s.$$

Now we are ready to state the theorem which gives bounds on root clustering of (11), assuming A_0 has roots inside the given root clustering region Ω_2.

Theorem 5. The perturbed system matrix $A_0 + E(q)$ has eigenvalues inside the given region Ω_2 of (6) if

$$|q_j| < \left[\left(\frac{b_s}{a_s}\right)^2 + \frac{\sigma_{\min}(Q)}{a_s}\right]^{1/2} - \left(\frac{b_s}{a_s}\right) = \mu_{2_{rs}},$$

where

$$b_s = \sigma_{\max}\left\{\left[c_{02m}\left(\sum (E_{iap})_m + \sum (A_{oip})_m\right)\right] + c_{11m}\left(\sum (P_{aei})_m\right) + c_{01}\left(\sum (P_{ei})_m\right)\right\},$$

$$a_s = r\sigma_{\max}[2c_{02m}(P_{ep})_m + c_{11m}(P_{ee})_m],$$

where P satisfies (7) and $\mu_{2_{rs}}$ denotes the perturbation bound for root clustering for region of degree 2 for structured uncertainty

Proof. Similar to the proof of Theorem 3.

4 *Illustrative example*

To illustrate the theory, consider a simple example with the plant matrix (see Abdul-Wahab, 1991)

$$A_0 = \begin{bmatrix} -4.3 & -0.4 \\ 0.2 & -3.4 \end{bmatrix},$$

with eigenvalues $\lambda_1 = -4.2$ and $\lambda_2 = -3.5$. Let us consider a circular root clustering region in the left half of the complex plane with the center at $\beta = -4.0$ and radius $r = 1.0$. Then the bound on the unstructured uncertainty $\sigma_{\max}(E)$ is given by equation (10). Carrying out the computations with $Q = I$, we get

$$\mu_{2_{ru}} = 0.0341.$$

That is, as long as the unstructured uncertainty is such that

$$\sigma_{\max}(E) < 0.0341,$$

the eigenvalues of $A_0 + E$ stay inside the circular region of the complex plane with the center at -4.0 and the radius $r = 1.0$.

5. *Conclusions*

This paper presented a unified theory for matrix root clustering for linear state space models (either in continuous time or in discrete time domain) subject to real parameter uncertainty. The method explicitly relates the root clustering region perameters to the parameter perturbation ranges for a

class of root clustering regions described by algebraic expressions. Since the method uses the Generalized Lyapunov Theory for getting the bounds the problem of conservatism is still present. Again the idea of improving the bounds by state transformation can be employed in this context also as was done for the case of robust stabilization in Yedavalli and Liang (1986). Efforts are underway to obtain non-conservative bounds for robust root clustering using Kronecker matrices and will be reported later.

Acknowledgements—This research is sponsored by NASA Langley Research Center under Grant # NAG-1-1164.

References

Abdul-Wahab, A. A. (1990). Lyapunov bounds for root clustering in the presence of system uncertainty. *Int. J. of Syst. Sci.*, **21**, 2603.

Abdul-Wahab, A. A. (1991). Perturbation bounds for root clustering of linear continuous time systems. *Int. J. of Syst. Sci.*, **22**, 921.

Ackermann, J., D. Kaesbauer and R. Muench (1991). Robust Gamma stability analysis in a plant parameter space. *Automatica*, **27**, 75.

Barmish, B. R. (1989). A generalization of Kharitonov's four-polynomial concept for robust stability problems with linearly dependent coefficient perturbations. *IEEE Trans. Aut. Control*, **AC-34**, 157.

Dorato, P. and R. K. Yedavalli (Eds) (1990). *Recent Advances in Robust Control.* IEEE Press, New York.

Fu, M. and B. R. Barmish (1989). Polytopes of polynomials with zeros in a prescribed set. *IEEE Trans. Aut. Control*, **AC-34**, 544.

Gutman, S. and E. I. Jury (1981). A general theory for matrix root clustering in subregions of the complex plane, *IEEE Trans. Aut. Control*, **AC-26**, 403.

Juang, Y. T. (1991). Robust stability and robust pole assignment of linear systems with structured uncertainty. *IEEE Trans. Aut. Control*, **AC-36**, 635.

Juang, Y. T., Z. C. Hong and Y. T. Wang (1989). Robustness of pole-assignment in a specified region. *IEEE Trans. Aut. Control*, **AC-34**, 758.

Keel, L. H., S. P. Bhattacharyya and J. W. Howze (1988). Robust control with structured perturbations. *IEEE Trans. Aut. Control*, **AC-33**, 68.

Keel, L. H., K. B. Lim and J. N. Juang (1991). Robust eigenvalue assignment with maximal tolerance to system uncertainties. *J. of Guidance, Control and Dynamics*, **14**, 615.

Kokame, H. and T. Mori (1991). A root distribution criterion for interval polynomials. *IEEE Trans. Aut. Control*, **AC-36**, 362.

Kolla, S. R., R. K. Yedavalli and J. B. Farison (1989) Robust stability bounds of linear discrete time systems. *Int. J. Contr.*, **50**, 151.

Patel, R. V. and M. Toda (1980). Quantitative measures of robustness for multivariable systems. *Proc. Joint Amer. Contr. Conf.*, TPS-A.

Siljak, D. D. (1989). Parameter space methods for robust control design: A guided tour. *IEEE Trans. Aut. Control*, **AC-34**, 674.

Sobel, K. M. and Y. Yu (1989). Flight control application of eigenstructure assignment with optimization of robustness to structured state space uncertainty. *Proc. 25th IEEE Conf. Decision Contr.*, Tempa, 1705.

Soh, C. B. (1989). On extending the hypersphere method to handle dominant pole assignment. *IEEE Trans. Aut. Control*, **AC-34**, 543.

Tesi, A. and A. Vicino (1990). Robust stability of state space models with structured uncertainties. *IEEE Trans. Aut. Control*, **AC-35**, 191.

Vicino, A. (1989). Robustness of pole location in perturbed systems. *Automatica*, **25**, 109.

Yedavalli, R. K. (1992). Counterexample to a recent result on perturbation bounds for root clustering by A. A. Abdul-Wahab. *Int. J. Syst. Sci.*, **23**, 661.

Yedavalli, R. K. and Z. Liang (1986). Reduced conservatism in stability robustness bounds by state transformation. *IEEE Trans. Aut. Control*, **AC-31**, 863.

Zeheb, E. (1989). Necessary and sufficient conditions for root clustering of a polytope of polynomials in a simply connected domain. *IEEE Trans. Aut. Control*, **AC-34**, 986.

Zhou, K. and P. Khargonekar (1987). Stability robustness bounds for linear state space models with structured uncertainty. *IEEE Trans. Aut. Control*, **AC-32**, 621.

Appendix A: Proof of Theorem 3

Let

$$\sigma_{\max}(E) < \frac{-d + (d^2 + 4ac)^{1/2}}{2a}, \tag{12}$$

where

$$a = 2c_{02m} + c_{11m}, \quad b = \sigma_{\max}(A_0),$$
$$c = \sigma_{\min}(Q)/\sigma_{\max}(P), \quad d = 4c_{02m}b + 2c_{11m}b + 2c_{01m}$$

and P satisfies

$$c_{00}P + c_{01}(PA_0^T + A_0P) + c_{11}A_0PA_0^T + c_{02}(PA_0^{T^2} + A_0^2P) = -Q.$$

Let

$$\sigma_{\max}(E) = x, \quad (.)_s = \frac{(.) + (.)^T}{2},$$

denote the symmetric part of matrix (.). Then (12) can be written as

$$x < \frac{-d + (d^2 + 4ac)^{1/2}}{2a},$$

$\Rightarrow$

$$ax^2 + dx - c < 0,$$

$\Rightarrow$

$$\begin{aligned}\sigma_{\max}(P)[2c_{01}\sigma_{\max}(E) + c_{11}\{2\sigma_{\max}(E)\sigma_{\max}(A_0)\\ + (\sigma_{\max}(E)^2\} + c_{02}\{2\sigma_{\max}(E)\sigma_{\max}(A_0)\\ + 2\sigma_{\max}(A_0)\sigma_{\max}(E) + 2(\sigma_{\max}(E)^2\}] < \sigma_{\min}(Q)\end{aligned}$$

$$\begin{aligned}\sigma_{\max}[2c_{01}(EP)_s + c_{11}(2(EPA_0^T)_s + EPE^T)\\ + c_{02}(2P(A_0E)_s + 2(EA_0)_sP + 2P(E^2)_s)] < \sigma_{\min}(Q)\end{aligned}$$

$\Rightarrow$

$$\begin{aligned}-Q + [2c_{01}(EP)_s + c_{11}(2(EPA_0^T)_s + EPE^T)\\ + c_{02}(2P(A_0E)_s + 2(EA_0)_sP + 2P(E^2)_s)] < 0\end{aligned}$$

$$\begin{aligned}c_{00}P + c_{01}[P(A_0 + E)^T + (A_0 + E)P]\\ + c_{11}[(A_0 + E)P(A_0 + E)^T]\\ + c_{02}[P((A_0 + E)^2)^T + (A_0 + E)^2P] < 0\end{aligned}$$

$\Rightarrow A_0 + E$ has eigenvalues inside the region Ω_2 of (6).

Automatica, Vol 29, No. 1, pp 241–244, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Necessary and Sufficient Conditions for Robust Stability of Discrete Systems with Coefficients Depending Continuously on Two Interval Parameters*

EZRA ZEHEB†

Key Words—Robust control; discrete systems; stability; zeros.

Abstract—Let a real polynomial in a complex variable, whose coefficients are any given continuous functions of two real interval parameters, be given.

Necessary and sufficient conditions are derived for the polynomial to have all its zeros outside (or inside) the unit circle of the complex variable plane.

In a multi-polynomial dependence of the coefficients on parameters, the conditions involve, in addition to a finite number of algebraic steps, only resolving single variable fixed coefficients real polynomials.

1. *Introduction*

IN ANALYSING AND DESIGNING realistic practical engineering systems, it is no longer sufficient to assume a "fixed" mathematical model of the system. Uncertainties about the model have to be taken into account. Indeed, ever since the publication of Kharitonov (1979) there has been an increasing interest in the subject. One approach to consider the uncertainties is through the coefficients of the characteristic polynomial of the system.

With regard to continuous-time systems, stability requires that the zeros of the characteristic polynomial be confined to the open left half complex plane. It was shown in Kharitonov (1979) that for the special case where the coefficients of the characteristic polynomial are independent "interval" parameters (i.e. they may take on any value within a given continuous interval) stability of a certain set of four polynomials is both necessary and sufficient to ensure stability of the entire family of polynomials. However, even for the simplest functional dependencies of coefficients on interval parameters, the situation becomes much more complicated. A major step towards the solution of the linear dependency case has been made in Bartlett *et al.* (1987). Three different solutions to this case were given in Fu and Barmish (1987), Barmish (1988) and Zeheb (1989). The multilinear dependency case has been assessed in Kraus *et al.* (1991a) and Barmish and Shi (1988). Necessary and sufficient conditions for the general case, where the coefficients of the characteristic polynomial are any continuous functions of the interval parameters (e.g. multipolynomial dependency) were given in Zeheb (1990).

With regard to discrete-time systems, stability requires that the zeros of the characteristic polynomial be confined to the exterior (or interior, depending on the definition) of the unit disk. Khartinov's result does not apply for this case (see e.g. Bose and Zeheb (1986)) even for the weaker theorem which requires stability of 2^n polynomials, where n is the number of coefficients (parameters) instead of four polynomials. However, some of the results for the linear dependency case in the methods of Barmish (1988) and Zeheb (1989) pertain to the unit disk as well. A review on robustness of discrete systems, along with a large list of relevant references, is given in Jury (1989).

Some other approaches to the problem and contributions towards its solution are given in Tsypkin *et al.* (1991), Anagnost *et al.* (1989), Kraus *et al.* (1991b), Qiu and Davison (1989), Chapellat and Bhattacharyya (1989), Kokame and Mori (1991), Anderson *et al.* (1987), Bartlett and Hollot (1988), Bose (1989), Ackermann and Barmish (1988), Bose *et al.* (1986) and others.

In fact, the solution to the general case, where the coefficients of the characteristic polynomial are any continuous functions of the parameters, can be treated by the D-decomposition graphical method (Neimark, 1947) presented in English in Lanzkron and Higgins (1959). This method employs mapping of the imaginary axis of the complex variable plane onto the plane of two variable parameters. Extensions make it useful for analysis of systems with more than two, but small number, of parameters. This method permits the designer to determine the number of the left half plane roots of the characteristic equation in various areas of the parameter plane, but the designer is unable to obtain information about, or have control over, the root locations. Another method Siljak (1966) which is graphical too, is also applicable to the above general case, by correlating the system parameters and the roots of the characteristic equation.

An analytic solution to the general case, where the coefficients of the characteristic polynomial are any continuous functions of the parameters, which is applicable to a general domain of root clustering, is derivable from Hertz *et al.* (1987), which is based on the zero sets concepts (Zaheb and Walach (1981)). The objective of the present paper is to *explicitly* formulate independent necessary and sufficient conditions for such a polynomial, whose coefficients are any given continuous functions of real interval parameters, to have all its zeros outside (or inside) the unit circle, and to provide direct and independent proofs to these conditions. For the sake of simplicity and explicitness, only two independent interval parameters are allowed. A preliminary version of the conditions (without proofs) were presented in Zeheb (1991). The computational complexity in testing these conditions for the multi-polynomial dependency case involves, in addition to a *finite* number of algebraic steps, only the resolution of single variable fixed coefficients real polynomials. However, it is not claimed that for a large number of parameters the computational complexity is manageable.

Since interval, or even polytopic, state space matrices of linear systems lead to characteristic polynomials with coefficients depending polynomially on the parameters, the

* Received 27 August 1991; received in final form 7 May 1992. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Guest Editor J. E. Ackermann.

† Department of Electrical Engineering, Technion Israel Institute of Technology, Haifa 32000, Israel.

results in the present paper also provide a solution to two of the research problems suggested in Jury (1991) namely:

(a) Obtaining necessary and sufficient conditions for robust stability of the classes of interval and polytope types of matrices for the discrete case.
(b) Study of robust Schur stability of a class of polynomials with coefficients depending multilinearly on perturbations.

2. *Necessary and sufficient conditions for robust Schur property in the continuous dependency case*

Let

$$Q(z, p_1, p_2) = \sum^{n} a_i z^i, \tag{1}$$

where

$$a_i = f_i(p_1, p_2), \quad i = 0, 1, \ldots, n, \quad a_i \in \mathbb{R}, \tag{2}$$

and f_i $(i = 0, \ldots, n)$ are continuous functions.

Theorem.

$$Q(z, p_1, p_2) \neq 0 \text{ in } |z| \leq 1 \text{ for } \underline{p}_i \leq p_i \leq \bar{p}_i, \quad i = 1, 2, \tag{3}$$

where $\underline{p}_i \leq \bar{p}_i$ are given real numbers, if, and only if,

(1) $Q(z, p_1^{(0)}, \bar{p}_2) \neq 0$ in $|z| \leq 1$ where $p_1^{(0)}$ is any arbitrary value in the interval $p_1^{(0)} \in [\underline{p}_1, \bar{p}_1]$.
(2) $Q(z, p_1, \bar{p}_2) \neq 0$ and $Q(z, p_1, \underline{p}_2) \neq 0$ in $|z| = 1$, $\underline{p}_1 \leq p_1 \leq \bar{p}_1$.
(3) $Q(z, \underline{p}_1, p_2) \neq 0$ and $Q(z, \bar{p}_1, p_2) \neq 0$ in $|z| = 1$, $\underline{p}_2 \leq p_2 \leq \bar{p}_2$.
(4) The following set of two real equations in two real unknowns

$$Q(1, p_1, p_2) = 0,$$
$$\frac{\partial Q(1, p_1, p_2)}{\partial p_1} = 0,$$

has no solution for which

$$\underline{p}_i \leq p_i \leq \bar{p}_i, \quad i = 1, 2.$$

(5) The following set of three real equations in three unknowns

$$Q(z, p_1, p_2) = \sum_{i=0}^{n} a_i z^i = 0,$$
$$z^n Q(z^{-1}, p_1, p_2) = \sum_{i=0}^{n} a_i z^{n-i} = 0,$$
$$\left(\sum_{i=0}^{n} \frac{\partial f_i}{\partial p_1} z^i\right)\left(\sum_{i=0}^{n} \frac{\partial f_i}{\partial p_2} z^{n-i}\right) - \left(\sum_{i=0}^{n} \frac{\partial f_i}{\partial p_1} z^{n-i}\right)\left(\sum_{i=0}^{n} \frac{\partial f_i}{\partial p_2} z^i\right) = 0,$$

has no solution for which

$$|z| = 1, \quad \underline{p}_i \leq p_i \leq \bar{p}_i, \quad i = 1, 2.$$

Remarks.

(1) Consider Condition 2. Note that for $|z| = 1$, z can be expressed as a complex function of a real parameter, e.g. $z = e^{j\theta}, 0 \leq \theta \leq 2\pi$. Hence, Condition 2 is equivalent to two sets of equations, each has two real equations (e.g. $\operatorname{Re} Q = 0$, $\operatorname{Im} Q = 0$ or $Q(z, p_1, \bar{p}_2) = 0$ and $Q(z^{-1}, p_1, \bar{p}_2) = 0$) in two real unknowns p_1 and θ, which must not have a solution for which $\underline{p}_1 \leq p_1 \leq \bar{p}_1, 0 \leq \theta \leq 2\pi$. A similar remark pertains to Conditions 3 and 5, where in the latter, the three unknowns become real for $|z| = 1$.
(2) The computational complexity involved in the above conditions is the following: Conditions 2–5 require the determination of whether a set of 2 (or 3) real equations in 2 (or 3) real unknowns has, or has not, a solution in given intervals. In a polynomial dependency case, one way to do it is to use the resultant method which, by a finite number of algebraic steps, reduces the problem to that of finding the zeros of a single variable fixed coefficients real polynomial in a given real interval. In many cases (where such zeros do not exist) it suffices to apply a Sturm test without having to solve the polynomial equation. In other cases, it is not recommended, from the computational viewpoint, to use the resultant method, especially in the three unknowns condition, because of introducing pseudo-solutions. Other algebraic and graphical methods to solve the problem may be more efficient, e.g. the Groebner Bases method or the Neimark D-decomposition method. Another possibility is to use the Jury–Marden table test (see e.g. Barnett (1983)) and require that the array is regular. The entries in the array are functions of p_1 (in Condition 2) or p_2 (in Condition 3) and their non-vanishing in the appropriate intervals can be checked by a Sturm test. Condition 1 can be carried out applying a Schur test (by one of the well-known criteria), which again amounts to a finite number of algebraic steps.
(3) If "stability" is defined by

$$Q(z, p_1, p_2) \neq 0 \text{ in } |z| \geq 1, \tag{4}$$

rather than by (3), the theorem is obviously applicable by changing variables $z \to 1/z$.

Proof of the Theorem. Necessity is obvious. We shall prove sufficiency. Consider the zero set V defined by

$$V = \{z \,.\, \exists p_1^0 \in [\underline{p}_1, \bar{p}_1], p_2^0 \in [\underline{p}_2, \bar{p}_2] \text{ such that } Q(z, p_1^0, p_2^0) = 0\}, \tag{5}$$

and denote the boundary of V by ∂V.

Obviously, (3) is equivalent to the requirement

$$(V) \cap (|z| \leq 1) = 0. \tag{6}$$

Now, if

$$(\partial V) \cap (|z| \leq 1) = 0, \tag{7}$$

then either (6) is satisfied or

$$(|z| \leq 1) \subset V. \tag{8}$$

Therefore, to satisfy (6), it suffices to satisfy (7) in addition to requiring $(z^0 = 1) \notin V$, i.e.

$$Q(1, p_1, p_2) \neq 0, \quad \underline{p}_1 \leq p_1 \leq \bar{p}_1, \quad \underline{p}_2 \leq p_2 \leq \bar{p}_2, \tag{9}$$

which contradicts (8).

Applying the "positivity theorem" in Walach and Zeheb (1980) to $Q(1, p_1, p_2)$ and to minus $Q(1, p_1, p_2)$ one obtains that (9) is satisfied iff Condition 4 is satisfied in addition to

$$Q(1, p_1, \bar{p}_2) \neq 0 \text{ and } Q(1, p_1, \underline{p}_2) \neq 0 \text{ for } \underline{p}_1 \leq p_1 \leq \bar{p}_1, \tag{9.1}$$

and

$$Q(1, \bar{p}_1, p_2) \neq 0 \text{ and } Q(1, \underline{p}_1, p_2) \neq 0 \text{ for } \underline{p}_2 \leq p_2 \leq \bar{p}_2. \tag{9.2}$$

Turn now to (7). If

$$(\partial V) \cap (|z| = 1) = 0, \tag{10}$$

then either (7) is satisfied or

$$(\partial V) \subset (|z| \leq 1). \tag{11}$$

By Theorem 1 in Zeheb and Walach (1981), a point $z^0 \in \partial V$ is either a point where one of the following four polynomials vanishes:

$$Q(z, p_1, \bar{p}_2), \quad Q(z, p_1, \underline{p}_2), \quad Q(z, \underline{p}_1, p_2), \quad Q(z, \bar{p}_1, p_2), \tag{12}$$

or else, z^0 is a solution of the set of equations

$$Q(z, p_1, p_2) = 0, \qquad \frac{\partial Q}{\partial p_1} \times \frac{\partial Q}{\partial p_2} = 0, \tag{13}$$

for which $\underline{p}_i \leq p_i \leq \bar{p}_i$, $i = 1, 2$.

The set of points complying with (12) and (13) is denoted by L, and obviously $\partial V \subset L \subset V$. Therefore, by requiring that $(L) \cap (|z| = 1) = 0$ and that at least one point of L does not belong to $|z| \le 1$, (7) is established. For $(L) \cap (|z| = 1) = 0$ Conditions 2–3 ensure that the four polynomials (12) do not vanish on $|z| = 1$ and, after some algebra and recalling that for $|z| = 1$, the complex conjugate of z equals $1/z$, Condition 5 ensures that (13) has no solution for $|z| = 1$.

Finally, Condition 1 ensures that at least one point of L does not belong to $|z| \le 1$.

Now since (9.1) and (9.2) are included in Conditions 2 and 3, Conditions 1–5 are sufficient to establish (3).

Remark. It beomes clear from the proof that the above necessary and sufficient conditions can be stated in different modified formulations. For example, in Condition 4 we may use the derivative $\partial Q(1, p_1, p_2)/\partial p_2 = 0$ instead of the derivative with regard to p_1 or, in Condition 1 we may use $Q(z, p_1^{(0)}, p_2) \neq 0$ in $|z| \le 1$. Another possibility is to use $(z^0 = 0) \notin V$ in (9). This would readily lead to the following conditions replacing Condition 4:

$$a_0(p_1, \bar{p}_2) > 0 \text{ and } a_0(p_1, \underline{p}_2) > 0 \text{ for } \underline{p}_1 \le p_1 \le \bar{p}_1, \quad (14)$$

$$a_0(\bar{p}_1, p_2) > 0 \text{ and } a_0(\underline{p}_1, p_2) > 0 \text{ for } \underline{p}_2 \le p_2 \le \bar{p}_2, \quad (15)$$

and the set of two real equations in two real unknowns

$$\begin{aligned} a_0(p_1, p_2) &= 0, \\ \frac{\partial a_0(p_1, p_2)}{\partial p_1} &= 0, \end{aligned} \quad (16)$$

has no solution for which

$$\underline{p}_i \le p_i \le \bar{p}_i, \quad i = 1, 2.$$

However, we will not elaborate on all, quite obvious, possibilities.

3. *A numerical example—polytopic state matrix*

Let a discrete time system be given by

$$x(n+1) = Ax(n), \quad (17)$$

where $x(\cdot)$ is the state vector and A, the system's matrix, has polytopic uncertainties, i.e.

$$A = A(p) = A_0 + A_1 p_1 + A_2 p_2,$$

$$A_0 = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$0 \le p_i \le 1, \quad i = 1, 2 \quad (18)$$

The characteristic polynomial becomes

$$\begin{aligned} Q(z, p_1, p_2) &= |zI - A| \\ &= z^3 - z^2(3 + 4p_1 + 3p_2) \\ &\quad + z(3 + 8p_1 + 6p_2 + 5p_1^2 + 3p_2^2 + 8p_1 p_2) \\ &\quad - (7 + 7p_1 + 14p_2 + 5p_1^2 + 9p_2^2 + 12p_1 p_2 \\ &\quad + 2p_1^3 + 2p_2^3 + 5p_1 p_2^2 + 5p_1^2 p_2). \end{aligned} \quad (19)$$

Note that $Q(z, p_1, p_2)$ is a third degree polynomial with coefficients depending *polynomially* (third degree) on two interval parameters.

According to our theorem,

$$Q(z, p_1, p_2) \neq 0 \text{ in } |z| \le 1 \text{ for } 0 \le p_i \le 1, \quad i = 1, 2,$$

if, and only if:

(1)

$$Q(z, p_1^{(0)} = 0, 1) = z^3 - 6z^2 + 12z - 32 \neq 0 \text{ in } |z| \le 1, \quad (20)$$

(20) is obviously satisfied since

$$32 > 1 + 6 + 12 = 19.$$

In a more general case, where the above sufficient condition is not relevant, a Schur test involving a finite number of algebraic steps may be carried out to determine the validity of this condition.

(2)

$$\begin{aligned} Q(z, p_1, 1) &= z^3 - z^2(6 + 4p_1) + z(12 + 16p_1 + 5p_1^2) \\ &\quad - (32 + 24p_1 + 10p_1^2 + 2p_1^3) \neq 0, \\ &\text{for } |z| = 1, \quad 0 \le p_1 \le 1. \end{aligned} \quad (21)$$

It can be checked almost by inspection that (21) is satisfied. For $|z| = 1$ it is sufficient that the coefficient of z^0, in absolute value, be greater than the sum of absolute values of the rest of the coefficients, i.e.

$$32 + 24p_1 + 10p_1^2 + 2p_1^3 > 1 + 6 + 4p_1 + 12 + 16p_1 + 5p_1^2,$$

or

$$2p_1^3 + 5p_1^2 + 4p_1 + 13 > 0,$$

which is always satisfied for $0 \le p_1 \le 1$. As indicated in Remark 2 above, a systematic algorithm can always be carried out in cases where the above sufficient condition is not satisfied

Turn now to the second part of Condition (2).

$$\begin{aligned} Q(z, p_1, 0) &= z^3 - z^2(3 + 4p_1) + z(3 + 8p_1 + 5p_1^2) \\ &\quad - (7 + 7p_1 + 5p_1^2 + 2p_1^3) \neq 0, \\ &\text{for } |z| = 1, \quad 0 \le p_1 \le 1, \end{aligned} \quad (22)$$

Using the Jury–Marden table form (see e.g. Barnett (1983)), it can be verified that (22) is also satisfied.

(3)

$$\begin{aligned} Q(z, 0, p_2) &= z^3 - z^2(3 + 3p_2) + z(3 + 6p_2 + 3p_2^2) \\ &\quad - (7 + 14p_2 + 9p_2^2 + 2p_2^3) \neq 0, \end{aligned} \quad (23)$$

for $|z| = 1$, $0 \le p_2 \le 1$.

As in the first part of Condition 2, the sufficient condition

$$7 + 14p_2 + 9p_2^2 + 2p_2^3 > 1 + 3 + 3p_2 + 3 + 6p_2 + 3p_2^2,$$

is obviously satisfied for $0 < p_2 \le 1$. For $p_2 = 0$, we obtain from (24)

$$z^3 - 3z^2 + 3z - 7 \neq 0 \text{ for } |z| = 1,$$

which is readily checked to be satisfied

The second part of Condition 3 is

$$\begin{aligned} Q(z, 1, p_2) &= z^3 - z^2(7 + 3p_2) + z(16 + 14p_2 + 3p_2^2) \\ &\quad - (21 + 31p_2 + 14p_2^2 + 2p_2^3) \neq 0, \end{aligned} \quad (24)$$

for $|z| = 1$, $0 \le p_2 \le 1$

Using again the Jury–Marden table form reveals that the coefficients are either all positive or all negative and thus (24) is also satisfied

(4)

$$\begin{aligned} Q(1, p_1, p_2) &\triangleq -2p_1^3 - 2p_2^3 - 5p_1 p_2^2 - 5p_1^2 p_2 - 6p_2^2 \\ &\quad - 4p_1 p_2 - 3p_1 - 11p_2 - 6 = 0, \end{aligned} \quad (25.1)$$

$$\frac{\partial Q(1, p_1, p_2)}{\partial p_1} \triangleq -6p_1^2 - 5p_2^2 - 10p_1 p_2 - 4p_2 - 3 = 0. \quad (25.2)$$

Obviously, *none* of these equations have a solution for which $0 \le p_i \le 1$, $i = 1, 2$ (and so does the set of equations) since all coefficients are negative.

(5) After some cumbersome but straightforward algebraic steps, the third equation in Condition 5 reaches the form

$$z[Az^4 + Bz^3 - Bz - A] = 0, \quad (26)$$

where

$$\begin{aligned} A &= 2p_1^2 + 9p_2^2 + 10p_1 p_2 + 18p_1 + 36p_2 + 35, \\ B &= -[2p_1^3 + 18p_2^3 + 40p_1 p_2^2 + 24p_1^2 p_2 \\ &\quad + 44p_1^2 + 90p_2^2 + 140p_1 p_2 \\ &\quad + 122p_1 + 142p_2 + 70]. \end{aligned}$$

The solutions of (26) are $z = 0$, $z = \pm 1$ and

$$Az^2 + Bz + A = 0. \tag{27}$$

the solutions $z = 0$ and $z = \pm 1$ are readily seen to be inconsistent with the other two equations in Condition 5.

Now, combining (27) with the other two equations in Condition 5, and since the coefficients of all three equations are real, the only possibilities of a common solution are:

(1) The left hand side of (27) is a factor of each of the two other equations.

(2) The common solution is real; i.e. for $|z| = 1$, it must be $z = 1$ or $z = -1$.

Both possibilities are readily checked to be inconsistent with $0 \le p_i \le 1$, $i = 1, 2$.

Hence, Condition 5 is satisfied, as well as the other four conditions of the Theorem. We conclude that the system (17) with polytopic state matrix uncertainties (18) is robustly stable.

4. *Conclusions*

Necessary and sufficient conditions are given, which solve the problem of testing for robust stability of a discrete system for the general case of *any continuous dependency* of the coefficients of the characteristic polynomial (or those of the state matrix) on two interval parameters.

The price for this generality is in computation complexity. However, at least for polynomial dependency (which includes, as special cases, the polytopic case and the multi-linear case) the number of required computational steps is *finite*, in reducing the problem into that of finding the roots of polynomials with fixed coefficients.

In many cases the computations may be simplified. It will be useful to investigate special cases where the computational complexity may be reduced.

Acknowledgement—This research was supported by the Fund for the Promotion of Research at the Technion.

References

Ackermann, J. E. and B. R. Barmish (1988). Robust Schur stability of a polytope of polynomials. *IEEE Trans. Aut. Contr.*, **33**, 984–986.

Anagnost, J. J., C. A. Desoer and R. J. Minnichelli (1989). Generalized Nyquist tests for robust stability: frequency domain generalizations of Kharitonov's theorem. In M. Milanese and R. Tempo (Ed.), *Robustness in Identification and Control* (*Conf. Proc.*), 79–96. Plenum Press, New York.

Anderson, B. D. O., E. I. Jury and M. Mansour (1987). On robust Hurwitz polynomials. *IEEE Trans. Aut. Contr.*, **32**, 909–913.

Barmish, B. R. (1988). A generalization of Kharitonov's four polynomial concept for robust stability problems with linearly dependent coefficient perturbations. *Proc. of the ACC*, Atlanta, GA.

Barmish, B. R. and Z. Shi (1988). Robust stability of a class of polynomials with coefficients depending multilinearly on perturbations. Rep. ECE 88-7, Univ. Wisconsin, Madison.

Barnett, S. (1983). *Polynomials and Linear Control Systems*. Marcel-Dekker, NY.

Bartlett, A. C. and C. V. Hollot (1988). A necessary and sufficient condition for Schur invariance and generalized stability of polytopes of polynomials. *IEEE Trans. Aut. Contr.*, **33**, 575–578.

Bartlett, A. C., C. V. Hollot and L. Huang (1987). Root locations of an entire polytope of polynomials: It suffices to check the edges. *Proc. ACC*.

Bose, N. K. (1989). Tests for Hurwitz and Schur properties of convex combinations of complex polynomials. *IEEE Trans. on Circ. and Syst.*, **36**, 1245–1247.

Bose, N. K. and E. Zeheb (1986). Kharitonov's theorem and stability test of multidimensional digital filters. *IEE Proc.*, **133**, 187–190.

Bose, N. K., E. I. Jury and E. Zeheb (1986). On robust Hurwitz and Schur polynomials. *Proc. 25th CDC*, Athens, pp. 739–744.

Chappellat, H. and S. P. Bhattacharyya (1989). A generalization of Kharitonov's theorem: robust stability of interval plants. *IEEE Trans. Aut. Contr.*, **34**, 306–311.

Fu, M. and B. R. Barmish (1987). Stability of convex and linear combinations of polynomials and matrices arising in robustness problems. *Proc. Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD.

Hertz, D., E. I. Jury and E. Zeheb (1987). Root exclusion from complex polydomains and some of its applications, *Automatica*, **23**, 399–404.

Jury, E. I. (1989). Robustness of discrete systems: A review, IFAC Working Group on Robust Control meeting, Jerusalem, Israel, March 1989.

Kharitonov, V. L. (1979). Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differential Equations*, **14**, 1483–1485.

Kokame, H. and T. Mori (1991). A Kharitonov-like theorem for interval polynomial matrices. *Systems and Control Letters*, **16**, 107–116.

Kaus, F. J., B. D. O. Anderson and M. Mansour (1991a). Robust stability of polynomials with multilinear parameter dependence, to be published.

Kraus, F. J., M. Mansour and E. I. Jury (1991b) Robust Schur stability of interval polynomials, to be published.

Lanzkron, R. W. and T. J. Higgins (1959). *D*-decomposition analysis of automatic control systems, *IRE Trans. Aut. Contr.*, **4**, 150–171.

Neimark, Yu. I. (1947). On the problem of the distribution of the roots of polynomials. *Dokl. Akad. Nauk SSSR*, **58**, 357–360.

Qiu, L. and E. J. Davison (1989). A simple procedure for the exact stability robustness computation of polynomials with affine coefficient perturbations. *Systems and Control Letters*, **13**, 413–420.

Siljak, D. D. (1966). Generalization of the parameter plane method. *IEEE Trans. Aut. Contr.*, **11**, 63–70

Tsypkin, Ya. Z., B. T. Polyak, A. Katbab and E. I. Jury (1991). On the strict and wide-sense stability robustness of uncertain systems—application of a new frequency criterion, private communication.

Walach, E. and E. Zeheb (1980). Sign test of multivariable real polynomials. *IEEE Trans. Circuits Syst.*, **27**, 619–625.

Zeheb, E. (1989). Necessary and sufficient conditions for root clustering of a polytope of polynomials in a simply connected domain. *IEEE Trans. Aut. Contr.*, **34**, 986–990.

Zeheb, E. (1990). Necessary and sufficient conditions for robust stability of a continuous system—The continuous dependency case illustrated via multilinear dependency. *IEEE Trans. Circ. and Syst.*, **CAS-37**, 47–53.

Zeheb, E. (1991). Necessary and sufficient conditions for robust Schur property—The continuous dependency case, presented in MTNS conference, Kobe, Japan.

Zeheb, E. and Walach (1981). Zero set of multiparameter functions and stability of multidimensional systems. *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-29**, 197–206.

Automatica, Vol. 29, No. 1, pp. 245–252, 1993
Pergamon Press Ltd. Printed in Great Britain.
International Federation of Automatic Control

Biographical Notes on Contributors to this Issue

B. Ross Barmish was born in Montreal, Quebec in 1949. He received the B.Sc. in Electrical Engineering from McGill University, Montreal, Quebec in 1971. In 1972 and 1975, respectively, he received the M.S. and Ph.D. degrees (both in Electrical Engineering) from Cornell University, Ithaca, NY. From 1975 to 1978 he served as Assistant Professor of Engineering and Applied Science at Yale University, New Haven, CT. From 1978 to 1984 he was an Associate Professor of Electrical Engineering at the University of Rochester, Rochester, NY. In August 1984 he joined the University of Wisconsin, Madison where he is currently Professor of Electrical and Computer Engineering. Dr Barmish has also been involved in a number of professional society activities; e.g. from 1982 to 1984, he served as an associate editor for the *IEEE Transactions on Automatic Control*; in 1983 and 1984, he also served on the IEEE Control Systems Society Board of Governors and the Technical Activities Board. Currently, he is an associate editor for the journal *Mathematics of Control, Signals and Systems*, a consulting editor for the *International Journal of Robust and Nonlinear Control* and the Program Chairman for the 1992 *American Control Conference*. His research interest lies in the robust control area and he is currently completing a textbook titled *New Tools for Robustness of Linear Systems* (to appear, Macmillan, New York).

Richard Y. Chiang was born in Taipei, Taiwan. He received a Ph.D. in Electrical Engineering from the University of Southern California in 1988. Between 1976 and 1985 he worked full time for various aerospace companies (CTCI, AiResearch, and Lear Astronics) as a control system analyst. From 1985 to 1988, he consulted numerous companies (STI, Northrop, General Dynamics Pomona, and Levitron Texas) while he was pursuing his Ph.D. degree. Between 1989 and 1991, he worked for the Flight Control Research Group of Northrop Aircraft. In 1991, he moved to the Guidance and Control Section of Jet Propulsion Lab. He is coauthor of the software package *Robust Control Toolbox* (1988, MathWorks Natick, MA.) and teaches part time at the University of Southern California. His current research interests include multivariable robust control theory, model reduction, system identification/control of large space structure, and attitude control of interplanetary spacecraft. He is a member of AIAA and IEEE.

Munther A. Dahleh was born in 1962. He received the B.S. degree from Texas A&M University, College Station, TX in 1983, and his Ph.D. degree from Rice University, Houston, TX, in 1987, all in Electrical Engineering. Since then, he has been with the Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, where he is now an Associate Professor. He has held consulting positions with C.S. Draper Laboratory and NASA since 1988. Professor Dahleh has received the Ralph Budd Award in Engineering in 1987, the George Axelby best paper award in 1989, a Presidential Young investigator award in 1991, and the Finmeccanica career development award in 1992. His current interests include robust control, identification of uncertain systems and adaptation, and control of time-varying and infinite-dimensional systems.

Dr Dahleh is currently serving as an Associate Editor for *Systems and Control Letters*.

Peter Dorato was born in New York, NY, 17 December 1932. He received his B.S.E.E. degree from the City college of New York in 1955, the M.S.E.E. degree from Columbia University in 1956, and the D.E.E. degree from the Polytechnic University (formally the Polytechnic Institute of Brooklyn) in 1961.

Dr Dorato was a faculty member at the City College of New York, from 1956 to 1957, the Polytechnic Institute of Brooklyn from 1961 to 1972, the University of Colorado at Colorado Springs, from 1972 to 1976, and the University of New Mexico from 1976 to the present time, where he is Professor of Electrical and Computer Engineering. He was chairman of the Electrical and Computer Engineering department from 1976 to 1984, and during the academic years 1969–1970, 1984–1985, and 1991–1992 he was on sabbatical leave at the University of Colorado at Boulder, the University of California at Santa Barbara, and the Politecnico di Torino, Italy, respectively. He is also an honorary professor at the Nanjing Aeronautical Institute in China. His areas of interest include robust control, multivariable feedback systems, and engineering education. He has been a past associate editor of the *IEEE Transactions on Automatic Control*, the *IEEE Transactions on Education*, and a past editor of the IEEE Press Selected Reprint Volume Series. He edited the volume *Robust Control* in that series and co-edited the volumes *Recent Advances in Robust Control* and *Advances in Adaptive Control*. He is the author (with L. Fortuna and G. Muscato) of the book, *Robust Control for Unstructured Perturbations—An Introduction* (1992 Springer-Verlag, Berlin). Dr Dorato is a Fellow of the IEEE and a Distinguished Member of the IEEE control Systems Society. He is currently an associate editor of *Automatica* and a registered professional engineer (state of Colorado).

John C. Doyle received the BS and MS degrees in electrical engineering from MIT in 1977, and the Ph.D. degree in Mathematics from UC Berkeley in 1984. He is a Professor of Electrical Engineering at Caltech and has been a consultant to Honeywell Systems and Research Center since 1976.

His theoretical research interests include modeling and control of uncertain and nonlinear systems, matrix perturbation problems, operator theoretic methods, and μ. His theoretical work has been applied throughout the aerospace industry and is gaining acceptance in the process control industry. Current application interests include flexible structures, chemical process control, flight control, and control of unsteady fluid flow and combustion. Additional academic interests include the impact of control on system design, the role of neoteny in personal and social evolution, modeling and control of acute and chronic human response to exercise, and feminist critical theory, especially in the philosophy of science.

Dr Doyle is the recipient of the Hickernell Award, the Eckman Award, the IEEE Control Systems Society Centennial Outstanding Young Engineer Award, and the Bernard Friedman Award. He is an NSF Presidential Young Investigator, an ONR Young Investigator, and has co-authored two IEEE Transactions Best Paper Award winners, one of which won the IEEE Baker Prize.

Giovanni Fiorio was born in Bioglio, Piedmont, Italy, on 11 August 1929. He received the degree in Industrial Engineering from the Politecnico di Torino in 1954. After a specialization course in Electrical Engineering at the same Politecnico, he received a university teaching qualification in Electrical Engineering in 1964.

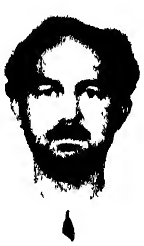
From 1954 to 1956 he was with an industrial society and successively he was teaching in a professional secondary school. From 1956 to 1969 he developed a research activity in the field of electrical measurements at the Istituto Elettrotecnico Nazionale Galileo Ferraris, in Turin, Italy. During this activity, he began his studies and researches in signal processing and automatic control. He is Full Professor at the Politecnico di Torino: Professor of Electrical Engineering from 1969 to 1982, and of Automatic Control from 1982 until now. His present research interests include modeling and identification of dynamic systems, process control, robustness properties of controllers with respect to different performance specifications.

Takao Fujii was born in Hiroshima, Japan in 1943. He received the B.E., M.E., and Ph.D. degrees, all in engineering, from Osaka University, Osaka, Japan, in 1967, 1969, and 1983, respectively.

From 1969 to 1988 he served as a Research Associate of Control Engineering at Osaka University. From 1988 to 1990 he was an Associate Professor of Mechanical Engineering for Computer-Controlled Machinery at Osaka University. In April 1990 he joined Kyushu Institute of Technology, Fukuoka, where he is currently a Professor of Control Engineering and Science. His research interests include linear quadratic optimal control and robust control theory and their applications.

He is an Associate Editor for the *Journal of Mathematical Systems, Estimation, and Control*. Dr Fujii is a member of the IEEE, the Society of Instrument and Control Engineering, and the Institute of Systems, Control, and Information Engineers.

Sanjay Garg was born in India in 1959. He received the Bachelor of Technology (B. Tech.) degree in Aeronautical Engineering from the Indian Institute of Technology, Kanpur, India in 1980, the M.Sc. degree in Aerospace Engineering from University of Minnesota in 1982, and the Ph.D. degree in Aeronautics and Astronautics from Purdue University in 1988. He worked as an Engineering Aide at the Systems Research Center, Honeywell Inc., Minneapolis while attending University of Minnesota. He has been working at NASA Lewis Research Center, Cleveland, OH, as a controls engineer since 1988, first as an employee of Sverdrup Technology Inc. and as a NASA employee since February 1991. He is the lead engineer on a program to develop an advanced Integrated Flight/Propulsion Control (IFPC) design methodology and demonstrate the methodology by application to Short Take-Off and Vertical Landing (STOVL) aircraft. His research interests include all aspects of application of modern multivariable robust control design techniques to aerospace vehicles. Dr Garg is a senior member of IEEE and AIAA and a member of the AIAA Technical Committee on Guidance, Navigation and Control



José C. Geromel was born in Itatiba, Brazil, in July 1952. He received the B.Sc. and M.Sc. degrees in electrical engineering from the State University of Campinas (UNICAMP), Campinas, Brazil, in 1975 and 1976, respectively, and the Docteur d'Etat es Sciences Phisiques degree from the University Paul Sabatier, Toulouse, France, in 1979. In 1975 he joined the Faculty of Electrical Engineering of the State University of Campinas where he is presently Professor of Control Theory and Systems Analysis. In 1987 he was a Visiting Professor in the Departimento di Elettronica of the Instituto Politecnico di Milano, Milan, Italy

Osvaldo Maria Grasselli was born in Rome, Italy, on 19 March 1942. He received the 'Laurea' degree in Electrical Engineering in 1968, and the post-graduate Diploma in Automatic Control in 1971, both from the University of Rome, Italy. In 1969 he joined the Istituto di Automatica of the University of Rome "La Sapienza" to work on multivariable control systems.

In 1971 he was secretary of the Italian Association of Automatic Control Researchers (G.R.A.). In 1972 he was

appointed researcher of the Centro di Studio dei Sistemi di Controllo e Calcolo Automatici of the National Research Council (C N R) Rome Italy

From 1973 to 1980 he was Associate Professor of System Theory at the University of Ancona Italy From 1980 to 1984 he was full Professor of System Theory and Dean of the Electrical Engineering College at the same University Since 1984 he is Full Professor of Control Theory at the Second University of Rome Tor Vergata where he was Dean of the Electrical Engineering College from 1985 to 1987

His main research interests are in the field of linear system and control theory

---

Isaac Kaminer received B S E E and M S E degrees from University Minnesota in 1983 and 1985, respectively From 1985 to 1989 he was with the Boeing Company where the worked in Guidance and Control Research Group for three years He is currently pursuing the Ph D degree at the University of Michigan EECS Department He expects to graduate in the summer of 1992 and join the faculty of Aerospace Department at Naval Postgraduate School in Monterey CA as an Assistant Professor

---

Da-Wei Gu graduated from the Department of Mathematics Fudan University Shanghai China in 1979 and received the M Sc degree in applied mathematics from Shanghai Jiao Tong University China in 1981 He was awarded the Ph D degree in control systems theory from the Department of Electrical Engineering Imperial College of Science and Technology London U K in 1985 During 1981–1982 he was a Lecturer in Shanghai Jiao Tong University From 1985 to 1989 he was a Research Assistant in the Department of Engineering Science Oxford University He has been a Lecturer in Control in the Engineering Department of Leicester University since 1989 His research interests include robust optimal control numerical optimization expert systems and software development for control system design

---

Hwan Il Kang was born in Inchon Republic of Korea on 27 November 1956 He received the B Sc in Electronics Engineering from Seoul National University Seoul Republic of Korea in 1980 and the M S degree in Electrical Science from Korea Advanced Institute of Science and Technology Seoul in 1982 He is currently a Ph D candidate in the department of Electrical and Computer Engineering at the University of Wisconsin Madison

From 1982 to 1984 he was a full time instructor of Electronics Engineering at the Kyungpook National University Daegu Republic of Korea He held a fellowship from the Korean government from 1984 to 1988 and held a research assistantship from January 1992 to October 1992 His current research interests include robustness of control systems and system optimization

---

Karl R Haiges received the B S degree in Electrical Engineering from Purdue University in 1981 and the M S degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 1984

Since 1984 he has been with the Northrop Corporation initially as a research engineer and now as a technical manager His research interests focus on the application of multivariable design and analysis methods to aircraft control laws

---

Ebrahim M Kasenally was born in London U K He received a B Sc (Eng) in Avionics from Queen Mary College (University of London) in 1984 and the M Sc and Ph D degrees in 1985 and 1989 respectively from Imperial College University of London He is currently a post doctoral research fellow in the Interdisciplinary Research Centre for Process Systems Engineering His research interests include large scale computations numerical linear algebra model reduction robust multivariable control and control of tokamaks

---

Dieter Kaesbauer was born in Landshut Germany in 1944 He received the M S degree in mathematics from the Technical University of Munich in 1970 and the Dr techn degree from the Technical University of Graz in 1986 Since 1971 he worked as a research scientist in the control group of the Institute for Flight Systems Dynamics at the German Aerospace Research Establishment (DLR) in Oberpfaffenhofen His research interests are in robustness analysis and control systems design

---

Mustafa Khammash was born in 1964 He received the B S degree in Electrical Engineering from Texas A&M College Station TX in 1986 and the Ph D degree in Electrical Engineering from Rice University Houston TX in 1990 In 1990 he joined the Electrical Engineering Dept at Iowa State University Ames IA where he is now an Assistant Professor Professor Khammash is the recipient of the Ralph Budd Award in Engineering His current research interests include robust and optimal control

Pramod P. Khargonekar received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay in 1977, the M.S. degree in mathematics and the Ph.D. degree in electrical engineering from the University of Florida in 1980 and 1981, respectively.

From 1981 to 1984, Dr Khargonekar was with the Department of Electrical Engineering, University of Florida, and from 1984 to 1989 he was with the Department of Electrical Engineering, University of Minnesota. In September 1989, he joined the University of Michigan where he holds the position of Professor of Electrical Engineering and Computer Science.

His current research interests include robust control, H_2, H_∞ and H_2/H_∞ optimal control, sampled-data systems, robust and H_∞ identification, robust adaptive control, time-varying systems, and applications to aerospace control problems.

Dr Khargonekar is a recipient of the American Automatic Control Council's Donald Eckman award, the NSF Presidential Young Investigator award, the George Taylor award from the University of Minnesota, and the Sigma Xi (University of Florida Chapter) outstanding research award. He is a co-recipient with Professors J. C. Doyle, B. A. Francis and K. Glover of the 1991 IEEE W. R. G. Baker Prize Award and the 1990 George Axelby Best Paper (in the *IEEE Transactions on Automatic Control*) award. He was an associate editor of the *IEEE Transactions on Automatic Control* during 1987–1989. He is currently an associate editor of *Mathematics of Control, Signals and Systems*, *SIAM Journal on Control and Optimization*, *Systems and Control Letters*, and *International Journal of Robust and Nonlinear Control*.

Jong-Lick Lin was born in Taiwan, Republic of China, in 1948. He received the B.S. degree and the M.S. degree in Electrical Engineering from the National Taiwan University, Taiwan, in 1973 and 1977, respectively. From 1977 to 1986, he was a Lecturer in the Department of Engineering Science at the National Cheng Kung University, Taiwan. In 1986 he became an Associate Professor in the same Department. In 1992 he completed a Ph.D. in the Engineering Department at Leicester University, U.K. working in the area of robust control.

Sauro Longhi was born in Loreto, Italy, on 11 September 1955. He received the Doctor degree in Electronic Engineering in 1979 from the University of Ancona, Italy, and the post-graduate Diploma in Automatic Control in 1985 from the University of Rome "La Sapienza", Italy. From 1980 to 1981 he held a fellowship at the University of Ancona. From 1981 to 1983 he was with the Telettra S.p.A., Chieti, Italy. Since 1983 he has been at the Dipartimento di Elettronica e Automatica of the University of Ancona, where he presently holds the position of Associate Professor. His main research interests concern analysis and control of linear systems.

David J. N. Limebeer was born in Johannesburg South Africa. He received the B.Sc.(Eng) degree in Electrical Engineering from the University of the Witwatersrand in 1974, and the M.Sc.(Eng) and Ph.D degrees in 1977 and 1980, respectively, from the University of Natal in South Africa. From 1974 to 1976 he was assistant engineer at the Johannesburg City Council. He was a Research Assistant at the University of Cambridge, U.K. between 1980 and 1983. In 1984 he moved to the Department of Electrical Engineering, Imperial College, London as a lecturer. In 1989 he was promoted to Reader in Control Engineering. In the summer of 1984 he had a summer position at the University of Southern California. He has held four Visiting Fellow appointments at the Australian National University, Canberra, Australia. He is currently an associate editor of *Automatica*, *Systems and Control Letters* and the *International Journal of Robust and Nonlinear Control*, and he has served on the British Science and Engineering Research Council's control and instrumentation sub-committee. He was joint recipient of the AACC O. Hugo Schuck award for the best paper presented at the 1983 ACC. His research interests include multivariable system theory, computer aided control system design, system identification, power system stability and the control of tokamaks. In 1991 he was elected fellow of the IEEE.

Kevin Madden was born in the U.S.A. He received a B.S. degree in Aerospace Engineering from University of Notre Dame in 1987 and an M.S. degree in Aerospace Engineering from California Polytechnic University, Pomona, CA in 1991. From 1987 to 1991 he worked with the Flight Control Research group of Northrop Aircraft, Hawthorne, CA. He is currently pursuing the MBA degree at Notre Dame University.

Stefano Malan was born in Torino, Italy, in 1963. He received the Laurea in electrical engineering from the Politecnico di Torino in 1988. From 1988 to 1990 he worked as a control engineer at FIAT Research Center. He is presently Researcher of Automatic Control at Dipartimento di Automatica e Informatica of the Politecnico di Torino. His research interests are mainly in the field of robust control theory.

Riccardo Marino was born in Ferrara, Italy, in 1956. He received the degree in nuclear engineering in 1979 and a Master degree in systems engineering in 1981, both from the university of Rome "La Sapienza" and the Doctor of Science degree in system science and mathematics from Washington University, St. Louis, MO, U.S.A., in 1982.

Since 1984 he has been with the Department of Electronic Engineering at the University of Rome "Tor Vergata", where he is currently a full professor in systems theory. He visited the University of Illinois at Urbana-Champaign during the academic years 1985–1986 and 1988–1989 and the University of Twente, Enschede, The Netherlands, in 1986. His research interest include theory and applications of nonlinear control and, more recently, nonlinear adaptive control.

Mario Milanese was born in Alessandria, Italy, on 11 September 1942. He received the Laurea Degree in Electronic Engineering from the Politecnico di Torino in 1967.

From 1967 to 1980 he was Assistant Professor at the Politecnico di Torino and from 1972 to 1981 Associate Professor of System Theory at the University of Turin. Since 1980 he has been Full Professor of System Theory at the Politecnico di Torino, where from 1982 to 1987 he has been Chairman of the Dipartimento di Automatica e Informatica.

He is Associate Editor of *Information and Decision Technologies* (North Holland, New York).

His research interests include modeling, robust identification and control, and set membership estimation theory.

Giovanni Muscato was born in Catania, Italy, in 1965. He received the electrical engineering degree from the University of Catania, in 1988. Following graduation he worked with "Centro di Studi sui Sistemi" in Turin. Since 1990 he joined the Dipartimento Elettrico Elettronico e Sistemistico of the University of Catania, where he presently holds the position of Assistant Professor of Automatic Control. His principal research interest include robust control, model reduction and the use of neural-networks in the modeling and control of dynamical systems.

Andy Packard is an assistant professor in Mechanical Engineering at the University of California, Berkeley. He earned his Ph.D. in Mechanical Engineering at UC Berkeley in January 1988. He was a research fellow and instructor at Caltech in 1988, and an assistant professor in electrical Engineering at University of California, Santa Barbara in 1989. He is an NSF Presidential Young Investigator, and his main technical interests are robust control and control system design.

John Perkins was born in the U.K. He received a B.Sc. (Eng) degree in Chemical Engineering and a Ph.D. degree from Imperial College London completing his studies there in 1973. From 1973 to 1977 he was University Demonstrator in Chemical Engineering at Cambridge University, with a year's leave of absence spent working at ICI Agricultural Division. In 1977, he returned to Imperial as a lecturer. He spent the years between 1985 and 1988 in Australia as the first holder of the ICI Chair in Process Systems Engineering at the University of Sydney, returning to a personal chair at Imperial College. He was elected a Fellow of the Institution of Chemical Engineers in 1986. He is Chief Editor of the *Journal of Process Control*, a member of the U.K. Science and Engineering Research Council's process engineering committee, and of the IFAC Working Group on Chemical Process Control.

In October 1992, he became director of the Interdisciplinary Research Centre in Process Systems Engineering at Imperial College in succession to Professor R. W. H. Sargent.

Pedro L. D. Peres was born in Sorocaba—SP, Brazil, in 1960. He received the B.Sc. and M.Sc. degrees in electrical engineering from the State University of Campinas, UNICAMP, in 1982 and 1985, respectively, and the "Doctorat en Automatique" degree from the University Paul Sabatier, Toulouse, France, in 1989. In 1990 he joined the Faculty of Electrical Engineering of the State University of Campinas, where he is currently a lecturer on Control and Optimization.

Ian Postlethwaite was born in Wigan, U.K. in 1953. He received the B.Sc. (Eng.) degree from Imperial College, London University in 1975 and the Ph.D. degree from Cambridge University in 1978.

From 1978 to 1981 he was a Research Fellow at Cambridge University and spent six months at General Electric Company, Schenectady, U.S.A. In 1981 he was appointed to a Lectureship in Engineering Science at Oxford University and to a Tutorial Fellowship at Oriel College. He was a Visiting Fellow at the Australian National University, Canberra from 1987 to 1988, whilst on sabbatical leave from Oxford. In 1988 he was appointed to a Chair of Engineering at Leicester University. His current research interests include robust control, optimal control and the application of advanced control system design to real engineering systems. He is a Member of the Institution of Electrical Engineers and a Member of the Institute of Electrical and Electronics Engineers.

Laura R. Ray is an Assistant Professor of Mechanical Engineering at Clemson University. She received her B.S.E. and Ph.D. degrees from the Department of Mechanical and Aerospace Engineering, Princeton University (1984, 1991) and her M.S. degree from the Department of Mechanical Engineering, Stanford University (1985). Her research interests include control theory, linear and nonlinear system robustness, state estimation, and applications of control theory to automotive, air transportation, and robotic systems. She has authored or co-authored several technical papers on robustness and control applications. She is a member of the AIAA and the ASME.

Mario A. Rotea was born in Rosario, Argentina, on 6 August 1958. He received the degree of Electronic Engineer from the National University of Rosario, Argentina in 1983. He obtained the M.S. degree in Electrical Engineering and the Ph.D. degree in Control Science and Dynamical Systems from the University of Minnesota in 1988 and 1990, respectively. In 1989 he was awarded a Doctoral Dissertation Fellowship by the Graduate School at the University of Minnesota.

From 1983 to 1984, he was an Assistant Engineer at the Military Ammunition Factory "Fray Luis Beltran", Argentina. From 1984 to 1986 he was a Research Associate at the Institute of Technological Development for the Chemical Industry, Santa Fe, Argentina. Currently, Dr Rotea is an Assistant Professor in the School of Aeronautics and Astronautics at Purdue University. His research interests include robust multivariable control, optimal control and applications of control theory to aerospace problems.

Michael G. Safonov was born in Pasadena, CA, on 1 November 1948. He received the B.S., M.S., Engineer, and Ph.D. degrees in Electrical Engineering from the Massachusetts Institute of Technology, Cambridge, MA in 1971, 1971, 1976 and 1977, respectively. From 1972 to 1975 he served with the U.S. Navy as Electronics Division Officer aboard the aircraft carrier Franklin D. Roosevelt (CVA-42). Since 1977 he has been with the University of Southern California where he is presently a Professor of Electrical Engineering and Associate Department Chairman. He has been a consultant to The Analytic Sciences Corp., Honeywell Systems and Research Center, Systems Control, Systems Control Technology, Scientific Systems, United Technologies, TRW, Northrop Aircraft, and Hughes Aircraft. During the academic year 1983–1984 he was a Senior Visiting Fellow with the Department of Engineering, Cambridge University, U.K., and in summer 1987 he held a similar appointment at Imperial College of Science and Technology, London, U.K. In 1990–1991 he was a visiting faculty member at Caltech, Pasadena, CA. He has authored or coauthored more than 100 journal and conference papers and the book *Stability and Robustness of Multivariable Feedback Systems* (1980, MIT Press, Cambridge, MA). Additionally, he is co-author of the *Robust-Control Toolbox* (1988, MathWorks, S. Natick, MA) a software package for use with PC-MATLAB. His research interests include robust control, infinity-norm optimal control theory and nonlinear system theory with applications to aerospace control design problems. He served as an Associate Editor of the *IEEE Trans. on Automatic Control* from 1985 to 1987 and is presently an editor of *International Journal of Robust and Nonlinear Control.* Dr Safonov is a Senior Member of the AIAA and a Fellow of the IEEE.

Jeff S. Shamma was born in New York, NY, in November 1963 and raised in Pensacola, FL. He received the Ph.D. degree in Mechanical Engineering in 1988 from the Massachusetts Institute of Technology. After a one year postdoctoral stay at MIT, he joined the University of Minnesota where he was an Assistant Professor of Electrical Engineering from 1989 to 1992. He then joined the University of Texas at Austin where he is currently an Assistant Professor of Aerospace Engineering. His main research interest is robust control for nonlinear, time-varying, and adaptive systems.

Robert Stengel is a Professor of Mechanical and Aerospace Engineering at Princeton University, where he directs the Topical Program on Robotics and Intelligent Systems and the Laboratory for Control and Automation. Prior to his 1977 Princeton appointment, he was with The Analytic Sciences Corporation, Charles Stark Draper Laboratory, U.S. Air Force, and National Aeronautics and Space Administration. A principal designer of the Project Apollo Lunar Module manual attitude control logic, he also contributed to the design of the Space Shuttle guidance and control system.

Dr Stengel received degrees from MIT (Aeronautics and Astronautics, S.B., 1960) and Princeton University (Aerospace and Mechanical Sciences, M.S.E., M.A., Ph.D., 1965, 1966, 1968). He is an Associate Fellow of the AIAA, a Senior Member of the IEEE, and a Member of the SAE Aerospace Control and Guidance Systems Committee. Professional positions include Associate Editor at Large of the *IEEE Transactions on Automatic Control*, member of the *Journal of Micromechanics and Microengineering* Editorial Board, and Member of the Program Council for the New Jersey Space Grant Consortium. He was vice-chairman of the congressional Aeronautical Advisory Committee and Chairman of the AACC Awards Committee, and he has served on numerous governmental advisory committees.

Dr Stengel's current research focuses on system dynamics, control, and machine intelligence. He teaches courses on control and estimation, aircraft dynamics, space flight engineering, and aerospace guidance. Dr Stengel wrote the book, *Stochastic Optimal Control: Theory and Application* (J. Wiley, 1986) and is writing a book on aircraft dynamics and control. He has authored or co-authored over 150 technical papers and reports.

Jonathan Tekawy was born in Indonesia, on 19 September 1962. He received the B.S. degree with *cum laude* in Aerospace Engineering from the California Polytechnic State University, Pomona, in 1984, the M.S. degree in Aeronautical and Astronautical Engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1987, and the Ph.D. degree in Mechanical Engineering from the University of California, Los Angeles, in 1990. From 1985 to 1986 he was a teaching assistant in the MIT Department of Aeronautical and Astronautical Engineering. From 1987 to 1990 he was with the Northrop Corporation, Aircraft Division, Hawthorne, CA. In 1990 he joined The Aerospace Corporation, El Segundo, CA, where he is currently a member of the technical staff in the Control Analysis Department. His research interests include robust multivariable control design, control of large space structures and perturbation techniques. He is a member of Phi Kappa Phi and Sigma Gamma Tau.

Robert Tempo was born in Cuorgné, Italy, in 1956. In 1980 he received the Laurea Degree in Electrical Engineering from Politecnico di Torino. In 1984 Roberto Tempo joined the CNR (National Research Council of Italy) at the research center CENS (Centro di Elaborazione Numerale dei Segnali) of Torino where he is currently Director of Research of Systems and Computer Science. His main fields of research include robust identification and control of uncertain systems and computational complexity. Presently, Dr Tempo is an associate editor of *Automatica*.

Patrizio Tomei was born in Rome, Italy, on 21 June 1954. He received the "dottore" degree in electrical engineering from the University of Rome, La Sapienza, in 1980.

From 1980 to 1984 he was in industry working on the implementation of real-time process controls. From 1984 to 1992 he has been working in the University of Rome "Tor Vergata". He is now associate professor at the University of Perugia. His research interests are in adaptive control, nonlinear control, and robotics.

Antonio Tornambè was born in Rome, Italy, on 19 March 1961. He received the 'Laurea' degree in electronic engineering from the University of Rome in 1985.

Since 1985 he has been at the Electronic Engineering Department of the Second University of Rome Tor Vergata. From 1987 to 1989 he has also been at the Fondazione Ugo Bordoni as a research engineer. From 1989 to 1992 he has been an associate researcher at the Second University of Rome. Tor Vergata and since 1992 he is an associate professor. His current interest include control and system theory, and robotics

Taro Tsujino was born in Osaka, Japan in 1965. He received the B.E. and M.E. degrees, all in engineering, from Osaka University, Osaka, Japan, in 1988 and 1990, respectively.

In April 1990 he joined Kyushu Institute of Technology, Fukuoka, where he is currently a Research Associate of Control Engineering and Science. His research interests include robust control theory and its applications.

Mr Tsujino is a member of the Japan Society of Mechanical Engineers, the Society of Instrument and Control Engineering, and the Institute of Systems, Control, and Information Engineers.

Dr Yedavalli received his B.E and M.E. degrees from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree from the School of Aeronautics and Astronautics of Purdue University in the dynamics and control area in 1981. He was an Assistant Professor at the Stevens Institute of Technology from 1981 to 1985 and an Associate Professor at the University of Toledo from 1985 to 1987. In September 1987, he joined the Department of Aeronautical and Astronautical Engineering at the Ohio State University, Columbus, Ohio where he is currently a tenured Associate Professor.

Dr Yedavalli's research interests include robustness and sensitivity issues in linear uncertain dynamical systems, model reduction, dynamics and control of flexible structures with applications to aircraft, spacecraft and robotics control.

Dr Yedavalli is an Associate Fellow of AIAA and a senior member of IEEE He served as a coeditor for an IEEE press book on "Recent Advances in Robust Control'. He presently serves as an Associate Editor for the *International Journal of Systems Science* and the *AIAA Journal of Guidance, Control and Dynamics*. He served as the Chairman of the Technical Program Committee for the 1991 AIAA Guidance, Navigation and Control Conference held in New Orleans, LA. He was a member of the team of instructors that taught an AIAA Professional Short Course on "Robust Multivariable Control: Theory and Practice", in August 1992. He was on the technical program committees for 1992 ACC and CCA, and serves as a society review chairman for AIAA for the 1993 ACC. He is preparing to offer a tutorial workshop on "Robust Control for State Space Systems," (along with Professor Leitmann) at the upcoming 1992 CDC. He serves as a reviewer for many journals, conferences and NSF and gave invited presentations on the above research topics. He organized and chaired and co-chaired sessions in conferences such as CDC, ACC and AIAA GNC. He is a member of the IFAC Working Group on 'Robust Control' and a member of the IEEE Working Group on Linear Multivariable Systems. He was a consultant to the Lawrence Livermore National Laboratory for two years.

Ezra Zeheb was born in Haifa, Israel. He received the degrees of B.Sc., M.Sc. and D.Sc. in Electrical Engineering from the Technion—Israel Institute of Technology. He is presently a Full Professor in the Department of Electrical Engineering at the Technion. Previously, he has been a Professor of Electrical Engineering at Tel-Aviv University, Tel-Aviv, Israel (1985–1987), a Director of Electronics Department at the Ministry of Defence, Israel (1970–1972), a member of the Technical Staff at Bell Telephone Laboratories (1968), and has held, in years in between, faculty positions at the Technion. He also served as a Consultant to the Ministry of Industry and Commerce, Israel (1972–1975 and 1977–1985). He has held visiting faculty positions at the University of California, Davis, CA, at the University of Missouri, Columbia, MO, and at Stevens Institute of Technology, Hoboken, NJ.

Dr Zeheb is a Fellow of the IEEE for contributions to the theory and design of robustly stable systems, and a member of the American Mathematical Society, Israel Mathematical Union, and the Sigma Xi Society. He is on the Editorial Board of the *Journal of Multidimensional Systems and Signal Processing*, and his research interests are in control, signal processing and filters, in which he authored or co-authored more than 60 papers in prestigious journals. Dr Zeheb has been the Chairman of the IEEE Israel Section (1986–1988 and 1990–1992), the Chairman of the IFAC Working Group on Robust Control (1987–1991), the Chairman of the 15th (1987) and 17th (1991) Conferences of Electrical and Electronics Engineers in Israel, and Chairman and Member of various international conference committees.

Rongze Zhao graduated from the Department of Mathematics and Mechanics of Beijing University, China, in 1968. After graduation he was a lecturer on Mathematics at the Hebei College of Geology and the Beijing College of Economy, both in China. He received Master degree in Mathematics from the University of Connecticut, in the U.S.A. in 1989. Currently, he is a Ph.D. student and a research assistant in the Program of Control Science and dynamical systems, at the University of Minnesota, MN U.S.A. His present research interests are in nonlinear system control and geometric control theory.

Automatica, Vol. 29, No. 2, p 253, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Editorial Changes

A YEAR AGO, in the March 1992 issue, several editorial staff changes were described. Now there is another major change: Dr H. Austin Spang III, who has been evaluating papers in all fields of Control Applications (including computer control, software, computer aided control system design, as well as implementation and operational evaluation) for over 24 years, ever since *Automatica* became the IFAC Journal, wishes to retire as an editor of *Automatica* at the end of 1993. Through these many years he has provided a valuable contribution to *Automatica* and IFAC by selecting well-written papers describing the application of control theory to real world problems, the ultimate goal of automatic control. His service and experience will be missed. However, we are fortunate in that he has found a very well qualified editor, who has been working with him as an Associate Editor for a number of years, to succeed him. This new editor is:

Professor Yaman Arkun
Georgia Institute of Technology
School of Chemical Engineering
778 Atlantic Drive
Atlanta, GA 30332-0100
U.S.A.

Therefore, *effective immediately*, papers in the areas of interest indicated above must be sent directly to Professor Arkun for evaluation and possible publication in *Automatica*. No papers should be sent to Dr Spang—except revised papers which were previously evaluated by him. Note that he will continue to evaluate only revised papers through 1993.

Meanwhile, the other *Automatica* Editors, K. J. Åström, P M. Larsen, W. S. Levine, A P. Sage, T. Başar, C. C. Hang, T. Söderström, and Deputy Editor-in-Chief H Kwakernaak will continue their efforts to obtain and to evaluate significant contributions in their respective areas of interest. Again, all of them, including retiring Editors P. C. Parks and H. Austin Spang III, are to be congratulated on the excellent work they have done and the results they have achieved.

To make the editorial efforts of the editors more efficient and rapid, authors submitting papers for possible publication in *Automatica* should do the following: (a) prepare their papers in the format described on the inside back cover of *Automatica*, (b) designate only one author to correspond with the Editor to whom five (or preferably six, for more rapid evaluation) copies of their paper are sent for evaluation, (c) send the Editor-in-Chief a copy of the paper and a copy of the letter sent to the Editor, and (d) send the Fax and email numbers (and the telephone number) of the corresponding author to the Editor and the Editor-in-Chief for rapid communication in case questions about the paper arise. Items (b), (c), and (d) are particularly important to avoid confusion and delay in processing submitted papers and tracking the progress of the reviews.

As a reminder, the designated, corresponding author of a paper having multiple authors should send five copies (preferably six) to the Editor whose area of interest is most closely related to the subject of the paper. It should be emphasized that to be considered for publication in *Automatica*, the paper must be sent to the Editor *whether or not it is also submitted to an IFAC meeting*. Also one copy should be sent to the Editor-in-Chief to be recorded in a central, global, file which is used for monitoring the progress of the evaluation procedures.

As the editors change so does the list of Associate Editors The *Automatica* Editorial Board, including the Editor-in-Chief in particular, are deeply grateful for the outstanding and essential service which they and their referees provide.

Another significant change must be noted: *Automatica* Deputy Editor-in-Chief, Hubert Kwakernaak, has been designated as the Liaison Officer with IFAC Meeting Organizers to assist them in selecting appropriate meeting papers for further review and possible publication in *Automatica*. In connection with this he is also serving as an Advisory Editor to help Guest Editors of Special Issues establish publication schedules and guidelines and to assist them in preparing the issues for publication according to *Automatica* standards

With these continuing changes and additions, we expect that *Automatica* will be better prepared to meet future developments in the field of automatic control and thereby provide increased service to IFAC and to the international control community.

George S Axelby
Editor-in-Chief
Automatica

Automatica, Vol 29, No 2, pp 255–273, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Tutorial Paper

Robust Control and $\mathscr{H}_\infty$-Optimization—Tutorial Paper*

HUIBERT KWAKERNAAK†

Robust control systems may successfully be designed by $\mathscr{H}_\infty$-optimization, in particular, by reformulating the design problem as a mixed sensitivity problem.

Key Words—$\mathscr{H}_\infty$-optimal control; robust control

Abstract—The paper presents a tutorial exposition of $\mathscr{H}_\infty$-optimal regulation theory, emphasizing the relevance of the mixed sensitivity problem for robust control system design

1 INTRODUCTION

THE INVESTIGATION OF $\mathscr{H}_\infty$-optimization of control systems began in 1979 with a conference paper by Zames (1979), who considered the minimization of the ∞-norm of the sensitivity function of a single-input–single-output linear feedback system. The work dealt with some of the basic questions of "classical" control theory, and immediately caught a great deal of attention. It was soon extended to more general problems, in particular when it was recognized that the approach allows dealing with robustness far more directly than other optimization methods.

The name "$\mathscr{H}_\infty$-optimization" is somewhat unfortunate. $\mathscr{H}_\infty$ is one member of the family of spaces introduced by the mathematician Hardy. It is the space of functions on the complex plane that are analytic and bounded in the right-half plane. The space plays an important role in the deeper mathematics needed to solve $\mathscr{H}_\infty$-optimal control problems.

This paper presents a tutorial exposition of the subject. The emphasis is on explaining the relevance of $\mathscr{H}_\infty$-optimization for control engineering. The paper presents few new results, and does not at all do justice to the extensive theoretical and mathematical literature on the subject. The presentation is limited to single-input–single-output (SISO) control systems. Many of the arguments carry over to the multi-input–multi-output case but their implementation is necessarily more complex.

We preview some of the contents. In Section 2 we use Zames' original minimum-sensitivity problem to introduce $\mathscr{H}_\infty$-optimization. Section 3 is devoted to a discussion of stability robustness. A well-known stability robustness criterion first proposed by Doyle (1979) demonstrates the relevance of the ∞-norm for robustness. Doyle's criterion in its original form has severe shortcomings, owing to the oversimplified representation of plant perturbations. In Section 4 it is explained how the criterion quite easily can be extended to a much more powerful result that applies to a very general perturbation model. In Section 5 this result is used for a perturbation model that for lack of a better name we refer to as numerator–denominator perturbations. It leads to the "mixed sensitivity" $\mathscr{H}_\infty$ stability robustness test. It is only slightly more complicated than Doyle's original test, and far less conservative for low-frequency perturbations.

Minimization of the mixed sensitivity criterion results in "optimal" robustness. In Section 6 the resulting mixed sensitivity problem is discussed in some detail. It is shown that it can be used not only for robustness optimization or robustness improvement, but also for design for performance. The design method based on the mixed sensitivity criterion features frequency response shaping, type k control and specified high-frequency roll-off, and direct control over the closed-loop bandwidth and time response by means of dominant pole placement.

To illustrate these features two design examples are included. In Section 7 a textbook example is discussed that is simple enough to be completely transparent. In Section 8 the application of the mixed sensitivity method to a benchmark example involving ship course control is described.

Sections 9–11 briefly review the theory needed to solve $\mathscr{H}_\infty$-optimal regulation problems. In Section 9 it is shown that the mixed sensitivity problem is a special case of the so-called "standard" $\mathscr{H}_\infty$-optimal regulation problem. In Section 10 the frequency domain solution of the standard problem is outlined, while Section 11 describes the main features of the state space solution.

* Received 6 February 1992; revised 6 July 1992; received in final form 23 August 1992. The original version of this paper was presented at the IFAC Symposium on Design Methods of Control Systems which was held in Zürich, Switzerland during September 1991. The Published Proceedings of this IFAC Meeting may be ordered from. Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Editor K. J. Åström.

† Systems and Control Group, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

We do not go into much detail, because the theory is not easy, and the development of algorithms that can be used unthinkingly for applications is best left to specialists.

2. SENSITIVITY

$\mathcal{H}_\infty$-optimization of control systems deals with the minimization of the peak value of certain closed-loop frequency response functions. To clarify this, consider by way of example the basic SISO feedback system of Fig. 1. The plant has transfer function P and the compensator has transfer function C. The signal v represents a disturbance acting on the system and z is the control system output. Then from the signal balance equation $\hat{z} = \hat{v} - PC\hat{z}$, with the circumflex denoting the Laplace transform, it follows that $\hat{z} = S\hat{v}$, where

$$S = \frac{1}{1+PC}, \tag{1}$$

is the sensitivity function of the feedback system. As the name implies, the sensitivity function characterizes the sensitivity of the control system output to disturbances. Ideally, $S = 0$.

The problem originally considered by Zames (1979, 1981) is that of finding a compensator C that makes the closed-loop system stable and minimizes the peak value of the sensitivity function. This peak value (see Fig. 2) is defined as

$$\|S\|_\infty = \max_{\omega \in \mathbb{R}} |S(j\omega)|, \tag{2}$$

where $\mathbb{R}$ denotes the set of real numbers. Because for some functions the peak value may not be assumed for any finite frequency, we replace the maximum here and in the following by the supremum or least upper bound, so that

$$\|S\|_\infty = \sup_{\omega \in \mathbb{R}} |S(j\omega)|. \tag{3}$$

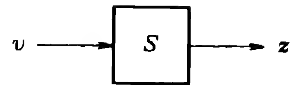


FIG. 3. Equivalent representation of the system of Fig. 1.

The justification of this problem is that if the peak value $\|S\|_\infty$ of the sensitivity function S is small, then the magnitude of S necessarily is small for all frequencies, so that disturbances are uniformly attenuated over all frequencies. Minimization of $\|S\|_\infty$ is worst-case optimization, because it amounts to minimizing the effect on the output of the worst disturbance (namely, a harmonic disturbance at the frequency where $|S|$ has its peak value).

The worst-case model has an important mathematical interpretation. Suppose that the disturbance v has unknown frequency content, but finite energy $\|v\|_2^2$. The number

$$\|v\|_2 = \sqrt{\int_{-\infty}^{\infty} |v(t)|^2 \, dt}, \tag{4}$$

is known as the 2-norm of the disturbance v. The energy of v is the square of the 2-norm. Then the norm $\|S\|$ of the system S as in Fig. 3 with input v and output z induced by the 2-norm is defined as

$$\|S\| = \sup_{v:\ \|v\|_2 < \infty} \frac{\|z\|_2}{\|v\|_2}. \tag{5}$$

Hence, in engineering terms the norm is directly related to the energy gain for the input with the worst possible frequency distribution. Using Parseval's theorem, it is not difficult to recognize that

$$\|S\| = \|S\|_\infty. \tag{6}$$

Hence, the peak value is precisely the norm of the system induced by the 2-norms on the input and output signals. This norm is known as the ∞-norm of the system.

It follows that $\mathcal{H}_\infty$-optimization is concerned with the minimization of system norms. It is useful to be aware of this when studying theoretical papers on $\mathcal{H}_\infty$-optimization.

Worst-case optimization suggests a game theory paradigm: The designer wishes to determine the compensator C that offers the best protection against the worst disturbance that nature has in store. This explains why in many theoretical papers $\mathcal{H}_\infty$-optimization is treated from the point of view of differential game theory.

A little contemplation reveals that minimization of $\|S\|_\infty$ as it stands is not a useful design tool. The frequency response function of every physical plant and compensator decreases at high frequencies. This means that often the sensitivity S can be made small at low frequencies but eventually reaches the asymptotic value one for high frequencies. Just how small S is at low frequencies is not reflected in the peak value but is of paramount importance for the control system

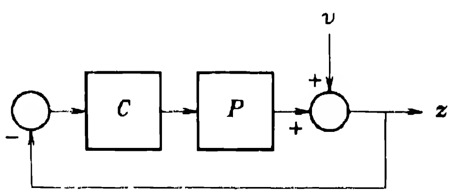


FIG. 1 SISO feedback loop.

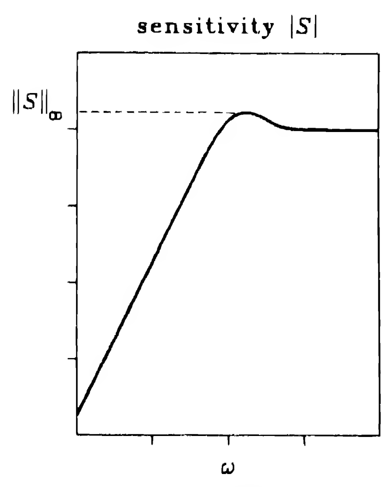


FIG. 2. $\|S\|_\infty$ as peak value.

performance For this reason, it is customary to introduce a frequency dependent weighting function W and consider the minimization of

$$\|WS\|_\infty = \sup_{\omega \in \mathbb{R}} |W(j\omega)S(j\omega)| \tag{7}$$

Characteristically, W is large at low frequencies but decreases at high frequencies

The weighted sensitivity minimization problem thus defined has interesting aspects Unfortunately, it does not account adequately for basic bandwidth limitations owing to restricted plant capacity, caused by the inability of the plant to absorb inputs that are too large Before going into this deeper we consider the question of robustness

3 ROBUSTNESS

We illustrate the connection between peak value minimization and design for robustness by considering in Fig 4 the Nyquist plot of the loop gain $L = PC$ of the SISO feedback system of Fig 1 In particular, we study whether the feedback system remains stable under a perturbation of the loop gain from its nominal value L_0 to the actual value L

For simplicity we take the system to be open loop stable (that is L represents a stable system) Naturally, we also assume that the nominal closed-loop system is well designed so that it is stable Then by the Nyquist stability criterion the Nyquist plot of the nominal loop gain L_0 does not encircle the point -1 The actual closed-loop system is stable if also the loop gain L does not encircle the point -1

It is easy to see by inspection of Fig 4 that the Nyquist plot L definitely does not encircle the point -1 if for every frequency ω the distance $|L(j\omega) - L_0(j\omega)|$ between any point $L(j\omega)$ on the plot of L and the corresponding point $L_0(j\omega)$ on the plot of L_0 is less than the distance $|L_0(j\omega) + 1|$ between the point $L_0(j\omega)$ and the point -1 that is if

$$|L(j\omega) - L_0(j\omega)| < |L_0(j\omega) + 1| \quad \text{for all } \omega \in \mathbb{R} \tag{8}$$

This is equivalent to

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \cdot \frac{|L_0(j\omega)|}{|L_0(j\omega) + 1|} < 1, \quad \text{for all } \omega \in \mathbb{R} \tag{9}$$

Define the complementary sensitivity function T_0 of the nominal closed-loop system as

$$T_0 = 1 - S_0 = 1 - \frac{1}{1 + L_0} = \frac{L_0}{1 + L_0}, \tag{10}$$

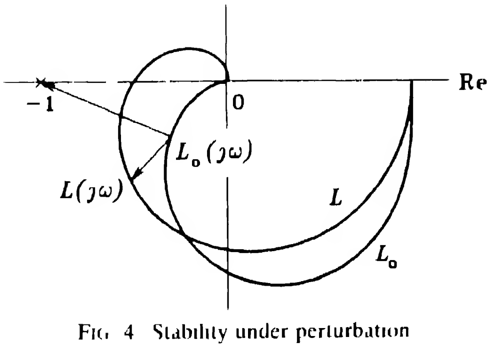


FIG 4 Stability under perturbation

with S_0 the nominal sensitivity function Then it follows from (9) that if

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \, |T_0(j\omega)| < 1, \quad \text{for all } \omega \in \mathbb{R}, \tag{11}$$

the perturbed closed-loop system is stable

The factor $|L(j\omega) - L_0(j\omega)|/|L_0(j\omega)|$ in this expression is the *relative* size of the perturbation of the loop gain L from its nominal value L_0 Suppose that this relative perturbation as a function of frequency is known to be bounded by

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} \le |W(j\omega)|, \quad \text{for all } \omega \in \mathbb{R}, \tag{12}$$

with W a given frequency dependent function Then

$$\frac{|L(j\omega) - L_0(j\omega)|}{|L_0(j\omega)|} |T_0(j\omega)|$$
$$= \frac{|L(j\omega) - L_0(j\omega)|/|L_0(j\omega)|}{|W(j\omega)|} |W(j\omega)T_0(j\omega)|$$
$$\le |W(j\omega)T_0(j\omega)| \tag{13}$$

Hence, if

$$|W(j\omega)T_0(j\omega)| < 1 \quad \text{for all } \omega \in \mathbb{R} \tag{14}$$

by (11) the closed-loop system is stable for all perturbations bounded by (12) Indeed, it may be shown that the condition (14) is not only sufficient but also necessary for the closed-loop system to be stable for all perturbations bounded by (12)

We obtained the condition (14) under the assumption that the open-loop system is stable It may be proved that it also holds for open-loop unstable systems, as long as the nominal and the perturbed open-loop system have the same number of right-half plane poles The result may also be extended to multivariable systems (Doyle 1979)

Using the norm notation introduced in (3) the condition (14) for robust stability may be rewritten as

$$\|WT_0\|_\infty < 1 \tag{15}$$

This explicitly demonstrates the relevance of the ∞-norm, that is, the peak value for robustness characterization The peak value criterion arises from the Nyquist stability criterion, which forbids the Nyquist plot of the loop gain to cross the point -1

For stability robustness the feedback system need be designed such that $\|WT_0\|_\infty$ is less than one It is tempting to consider the problem of minimizing the norm $\|WT_0\|$ with respect to all compensators that stabilize the closed-loop system as a way of optimizing robustness Stability seldom is the sole design target, though and robustness optimization may easily lead to useless results If the system is open-loop stable, for instance and all perturbations that may arise leave it stable $\|WT_0\|_\infty$ may be made equal to zero, and, hence, minimal, by simply letting $C = 0$, so that also $L_0 = 0$ and $T_0 = 0$ This optimizes stability robustness, but does nothing to improve control system performance, such as its sensitivity and response properties In Section 5 we introduce an alternative stability robustness criterion that allows consideration of the response properties as well

It is important to note that although it looks very plausible to characterize the plant perturbations by the bound (12), the bound may in fact allow far more perturbations than may actually occur. By way of example, suppose that the perturbations are caused by variation of a single parameter. Then if these variations have an important effect on the loop gain it may be necessary to choose W quite large to satisfy the bound. This large bound also allows many other perturbations, however, including such that change the order of the plant. Because of this the robustness stability test (14) may easily fail even if the actual parameter perturbations do not destabilize the feedback system.

This phenomenon, usually referred to as conservativeness, seriously handicaps the applicability of the robustness stability analysis described in this section. The model is mainly suited to deal with high-frequency uncertainty caused by parasitic effects and unmodelled dynamics that cause the envelope defined by (12) to be densely filled.

4. A GENERAL PERTURBATION MODEL

In the preceding section we considered perturbations of the form $L_0 \to L$, where $|L(j\omega) - L_0(j\omega)|/|L_0(j\omega)| \le |W(j\omega)|$ for all $\omega \in \mathbb{R}$. Equivalently, we may write

$$L_0 \to L_0(1 + \delta_L W), \tag{16}$$

where δ_L is any frequency dependent function such that $|\delta_L(j\omega)| \le 1$ for all $\omega \in \mathbb{R}$, that is, such that

$$\|\delta_L\|_\infty \le 1. \tag{17}$$

This perturbation may be represented as in the block diagram of Fig. 5. Because at this point we are only interested in stability the disturbance has been omitted. The functions W and δ_L are represented as frequency response functions of stable systems.

Figure 5 is a special case of the configuration of Fig. 6, where H represents the dashed block in Fig. 5. In the block diagram of Fig. 6 the perturbation δ is isolated from the rest of the system H.

By the small gain theorem (see for instance Desoer and Vidyasagar, 1975) a sufficient condition for the closed-loop system of Fig. 6 to be stable is that the norm $\|H\delta\|$ of the loop map $H\delta$ be less than 1. By

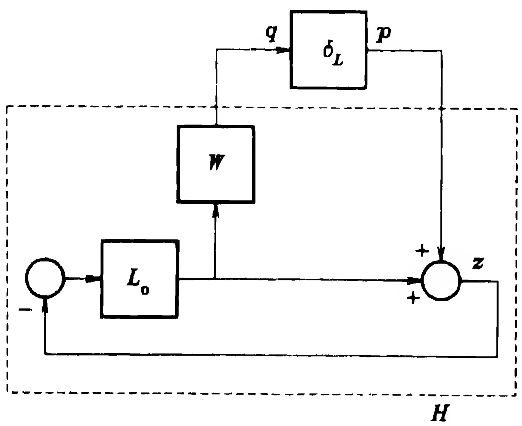


FIG. 5. Feeback loop with perturbation.

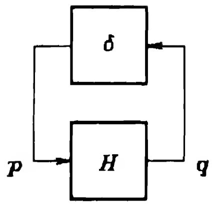


FIG. 6. General perturbation model.

the inequality $\|H\delta\| \le \|H\| \cdot \|\delta\|$ this is guaranteed if $\|H\| \cdot \|\delta\| < 1$. Taking in particular the ∞-norm it follows that the perturbed system is stable for all perturbations δ whose ∞-norm is, at most 1 if

$$\|H\|_\infty < 1. \tag{18}$$

The perturbation model of Fig. 6 with the corresponding condition (18) for stability with respect to all perturbations such that $\|\delta\|_\infty \le 1$ is simple yet very general. The condition (18) is not only sufficient but also necessary.

To illustrate its application we specialize the result to the configuration of Fig. 5. H is the transfer function from p to q after opening the loop by removing the block "δ". Inspection of Fig. 5 shows that $\hat{z} = \hat{p} - L_0\hat{z}$, so that

$$\hat{z} = \frac{1}{1 + L_0}\hat{p}. \tag{19}$$

By further inspection we have

$$\hat{q} = -WL_0\hat{z} = -W\frac{L_0}{1 + L_0}\hat{p} = -WT_0\hat{p}, \tag{20}$$

so that the transfer function H is given by

$$H = -WT_0. \tag{21}$$

Hence, by (18) stability under perturbation is guaranteed if

$$\|WT_0\|_\infty < 1. \tag{22}$$

This is precisely the stability robustness criterion of Section 3.

The general model of Fig. 6 with the necessary and sufficient condition $\|H\|_\infty < 1$ for robust stability with respect to all perturbations such that $\|\delta\|_\infty \le 1$ applies to SISO as well as MIMO systems. It was conceived by Doyle (1984). To use the condition for MIMO systems we need discuss how the ∞-norm is defined for such systems. Consider a stable MIMO system with input u, output y and transfer matrix F as in Fig. 7. The ∞-norm of the system is the norm induced by the norms

$$\begin{aligned} \|u\|_2 &= \sqrt{\int_{-\infty}^{\infty} u^H(t)u(t)\,dt}, \\ \|y\|_2 &= \sqrt{\int_{-\infty}^{\infty} y^H(t)y(t)\,dt}, \end{aligned} \tag{23}$$

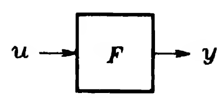


FIG. 7. MIMO system.

of the input and output, respectively. The superscript ^{H} denotes the complex conjugate transpose. Again using Parseval's theorem, it may be shown that the norm of the system induced by these signal norms is

$$\|F\|_\infty = \sup_{\omega \in \mathbb{R}} \|F(j\omega)\|_2, \tag{24}$$

where for a constant complex-valued matrix A the notation $\|A\|_2$ indicates the spectral norm

$$\|A\|_2 = \max_i \sigma_i(A), \tag{25}$$

with σ_i the ith singular value†. It follows from (24)–(25) that the ∞-norm $\|F\|_\infty$ of the stable system with transfer matrix F is found by first computing for each frequency the largest singular value of the frequency response matrix $F(j\omega)$, and then taking the maximum of all these largest singular values over frequency.

In the following section we discuss a further application of the basic stability robustness result.

5. NUMERATOR–DENOMINATOR PERTURBATIONS

In this section we discuss the application of the general stability robustness result to what we call numerator–denominator perturbations, or coprime factor uncertainty, as they are also known. The model may be traced to Vidyasagar (Vidyasagar *et al.*, 1982; Vidyasagar, 1985) and Kwakernaak (1983, 1986). It relies on the representation of the plant transfer function P in the block diagram of Fig. 1 in fractional form as

$$P = \frac{N}{D}. \tag{26}$$

In particular, if P is rational, N obviously can be taken as the numerator polynomial and D as the denominator polynomial. It is not necessary to do this, however, and indeed it sometimes is useful to arrange N and D differently. The numerator-denominator perturbation model represents perturbations in the form

$$P_0 = \frac{N_0}{D_0} \rightarrow P = \frac{N_0 + M\delta_N W_2}{D_0 + M\delta_D W_1}, \tag{27}$$

where the subscripts on N and D denote the nominal system. The terms $M\delta_D W_1$ and $M\delta_N W_2$ model the uncertainty in the denominator and the numerator respectively. The frequency dependent functions MW_1 and MW_2 represent the largest possible perturbations of the denominator and numerator, respectively, and δ_D and δ_N are frequency dependent functions of magnitude not greater than one. The factor M is included for added flexibility. Its use (for partial pole placement) becomes clear in Section 6.

† If A is an $n \times m$ matrix, the singular values of A are the $\min(n, m)$ largest of the m nonnegative numbers $\sqrt{\lambda_i(A^H A)}$, $i = 1, 2, \ldots, m$. Here λ_i denotes the ith eigenvalue. The singular values are also the $\min(n, m)$ largest of the n numbers $\sqrt{\lambda_i(AA^H)}$, $i = 1, 2, \ldots, n$.

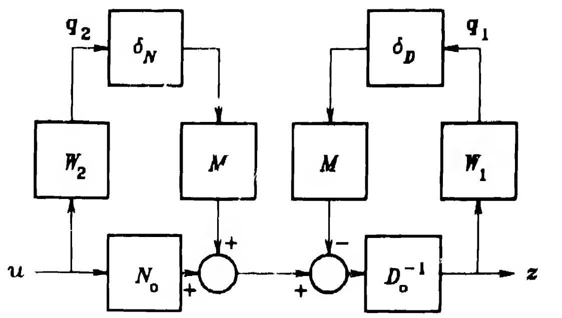


FIG. 8 Numerator–denominator perturbation model.

It is not difficult to check that the perturbation (27) may be represented as in the block diagram of Fig. 8. Note that the perturbation δ_D appears in a feedback loop. Including the plant in the control system configuration of Fig. 1 the block diagram may be arranged as in Fig. 9, where the block "δ_P" is described by

$$\hat p = -\delta_D \hat q_1 + \delta_N \hat q_2 = \underbrace{[-\delta_D \quad \delta_N]}_{\delta_P} \underbrace{\begin{bmatrix} \hat q_1 \\ \hat q_2 \end{bmatrix}}_{\hat q}. \tag{28}$$

The dashed lines in Fig. 9 indicate how the situation may be reduced to the perturbation model of Fig. 6. For the application of condition (18) for robust stability we need to assume that the closed-loop system is stable, and that moreover M, W_1, and W_2 have all their poles in the open left-half plane.

The block marked "H" has input p and output $q = \text{col}(q_1, q_2)$. To find the transfer matrix H, we first inspect Fig. 9 to establish the signal balance equation $\hat z = D_0^{-1}(M\hat p - N_0 C\hat z)$. It follows that

$$\hat z = \frac{D_0^{-1}M}{1 + D_0^{-1}N_0 C}\hat p = \frac{V}{1 + P_0 C}\hat p, \tag{29}$$

where we define $V = D_0^{-1}M$. By further inspection we see that

$$\begin{aligned} \hat q_1 &= W_1 \hat z = \frac{W_1 V}{1 + P_0 C}\hat p = W_1 S_0 V \hat p, \\ \hat q_2 &= -W_2 C \hat z = -\frac{W_2 C V}{1 + P_0 C}\hat p = -W_2 U_0 V \hat p, \end{aligned} \tag{30}$$

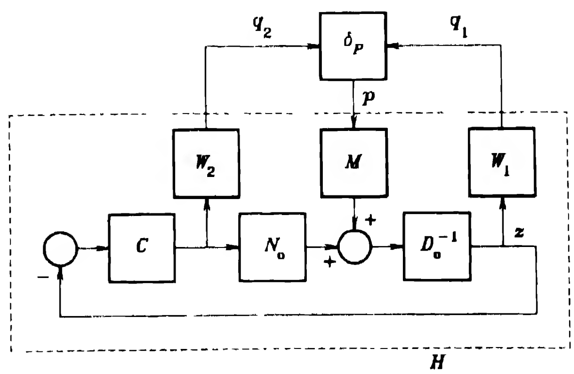


FIG. 9 Numerator–denominator perturbation model in feedback loop.

where
$$S_0 = \frac{1}{1+P_0C}, \quad U_0 = \frac{C}{1+P_0C}, \tag{31}$$

are the nominal sensitivity function and the nominal input sensitivity function, respectively, of the feedback system. The input sensitivity function U is the transfer function from the disturbance to the plant input. It is related to the complementary sensitivity function T as $T = PU$.

From (30) we see that the transfer matrix H in Fig. 9 is given by
$$H = \begin{bmatrix} W_1 S_0 V \\ -W_2 U_0 V \end{bmatrix}. \tag{32}$$

It is easy to check that the 1×2 matrix $A = [A_1 \ A_2]$ has the single singular value $\sqrt{|A_1|^2 + |A_2|^2}$. Hence, the square of the ∞-norm of the perturbation $\delta_P = [-\delta_D \ \delta_N]$ is
$$\|\delta\|_\infty^2 = \sup_{\omega \in \mathbb{R}} (|\delta_D(j\omega)|^2 + |\delta_N(j\omega)|^2). \tag{33}$$

The 2×1 matrix $A = [A_1 \ A_2]^T$ also has the single singular value $\sqrt{|A_1|^2 + |A_2|^2}$. Hence, the square of the ∞-norm of the system with transfer matrix H as given by (32) is
$$\|H\|_\infty^2 = \sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S_0(j\omega)V(j\omega)|^2 + |W_2(j\omega)U_0(j\omega)V(j\omega)|^2). \tag{34}$$

It follows that the closed-loop system is stable for all numerator–denominator perturbations $\delta_P = [-\delta_D \ \delta_N]$ satisfying the bound
$$|\delta_D(j\omega)|^2 + |\delta_N(j\omega)|^2 \le 1, \quad \text{for all } \omega \in \mathbb{R}, \tag{35}$$

if and only if the sensitivity function S_0 and the input sensitivity function U_0 satisfy the inequality
$$\|W_1(j\omega)S_0(j\omega)V(j\omega)|^2 + |W_2(j\omega)U_0(j\omega)V(j\omega)\|^2 < 1, \tag{36}$$
for all $\omega \in \mathbb{R}$.

Alternative interpretation

It is useful to consider an alternative interpretation of this stability robustness result. For the numerator–denominator perturbation model (27) the relative perturbation of the denominator is given by
$$\frac{D - D_0}{D_0} = \frac{M\delta_D W_1}{D_0} = VW_1\delta_D = w_1\delta_D, \tag{37}$$

where $w_1 = VW_1$. Similarly, the relative perturbation of the numerator is
$$\frac{N - N_0}{N_0} = \frac{M\delta_N W_2}{N_0} = \frac{VW_2}{P_0}\delta_N = w_2\delta_N, \tag{38}$$

where $w_2 = W_2V/P_0$. From (35) it follows with (37)–(38) that we consider perturbations satisfying
$$\left|\frac{\frac{D - D_0}{D_0}}{w_1}\right|^2 + \left|\frac{\frac{N - N_0}{N_0}}{w_2}\right|^2 \le 1, \tag{39}$$

on the imaginary axis. Substituting $W_1V = w_1$ and $W_2V = w_2P_0$ into (36) we see that the system is robustly stable for such perturbations if and only if
$$|S_0w_1|^2 + |T_0w_2|^2 < 1, \text{on the imaginary axis}, \tag{40}$$

since $P_0U_0 = T_0$.

By (39), the functions w_1 and w_2 are measures for the relative sizes of the perturbations in the denominator and the numerator of the plant transfer matrix P, respectively. The stability robustness test (40) then indicates that the nominal sensitivity function S_0 should be small for those frequencies where the relative perturbations in the denominator are large, and that the nominal complementary sensitivity function T_0 should be small for those frequencies where the relative perturbations in the numerator are large.

Conversely, this interpretation provides us with an indication how to model perturbations compatibly with performance requirements. Customary performance specifications require the sensitivity function to be small at low frequencies, and to level off to one at high frequencies (see Fig. 10). Designing the system

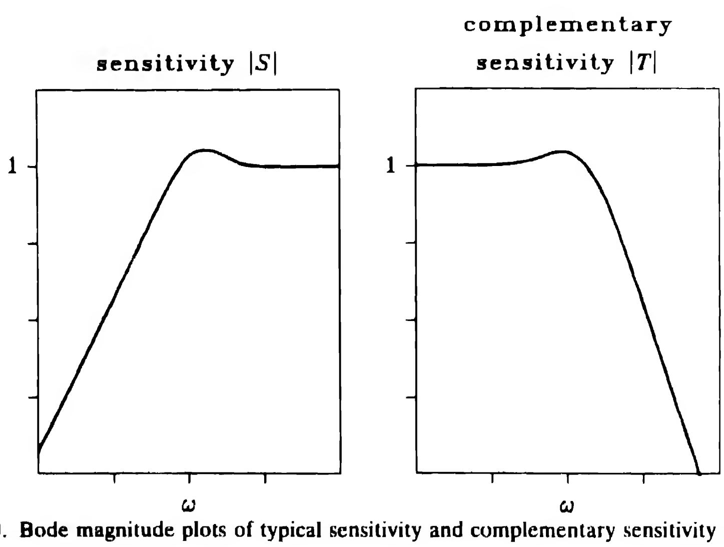


FIG. 10. Bode magnitude plots of typical sensitivity and complementary sensitivity functions.

this way ensures disturbance attentuation over the widest possible frequency range given the plant capacity (that is, given the largest inputs the plant can absorb).

This means that low-frequency perturbations are best modeled as denominator perturbations. The most important low-frequency perturbations are normally caused by parameter uncertainty, often referred to as structured uncertainty. On the other hand, high-frequency perturbations are best modeled as numerator perturbations, because (complementarily—see Fig. 10) the complementary sensitivity function T_0 is small at high frequencies. High-frequency perturbations are usually caused by parasitic effects and unmodeled dynamics, often known as unstructured uncertainty.

The examples in Sections 7 and 8 illustrate the application of these ideas.

6. THE MIXED SENSITIVITY PROBLEM

In Section 5 it was found that stability robustness with respect to numerator–denominator perturbations such that

$$|\delta_D(j\omega)|^2 + |\delta_N(j\omega)|^2 < 1, \quad \omega \in \mathbb{R}, \tag{41}$$

is guaranteed if $\|H\|_\infty < 1$, where (omitting the subscripts on S and U)

$$\|H\|_\infty^2 = \sup_{\omega \in \mathbb{R}} (|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2) \tag{42}$$

Given a feedback system with compensator C that does not satisfy this inequality one may look for a different compensator that does achieve inequality. An effective way of doing this is to consider the problem of minimizing $\|H\|_\infty$ with respect to all compensators C that stabilize the system. If the minimal value of $\|H\|_\infty$ is greater than 1, no compensator exists that stabilizes the systems for all perturbations satisfying (41). In this case, stability robustness is only obtained for perturbations satisfying (41) with the right-hand side replaced with $1/\lambda^2$.

The problem of minimizing

$$\left\| \begin{bmatrix} W_1SV \\ W_2UV \end{bmatrix} \right\|_\infty, \tag{43}$$

(Kwakernaak, 1983, 1985) is a version of what is known as the mixed sensitivity problem (Verma and Jonckheere, 1984). The name derives from the fact that the optimization involves both the sensitivity and the input sensitivity function (or, in other versions, the complementary sensitivity function).

In what follows we explain that the mixed sensitivity problem cannot only be used to verify stability robustness for a class of perturbations, but also to achieve a number of important design targets for the one-degree-of-freedom feedback configuration of Fig. 1.

Frequency response shaping

The mixed sensitivity problem may be used for performance design by shaping the sensitivity and input sensitivity functions. The reason is that the solution of the mixed sensitivity problem has the property that the frequency dependent function

$$|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2, \tag{44}$$

whose peak value is minimized, actually is a constant (Kwakernaak, 1985). This is known as the equalizing property. If we denote the constant as λ^2, with λ nonnegative, it immediately follows from

$$|W_1(j\omega)S(j\omega)V(j\omega)|^2 + |W_2(j\omega)U(j\omega)V(j\omega)|^2 = \lambda^2, \tag{45}$$

that for the optimal solution

$$\begin{aligned} |W_1(j\omega)S(j\omega)V(j\omega)|^2 &\le \lambda^2, \quad \omega \in \mathbb{R}, \\ |W_2(j\omega)U(j\omega)V(j\omega)|^2 &\le \lambda^2, \quad \omega \in \mathbb{R}. \end{aligned} \tag{46}$$

Hence,

$$|S(j\omega)| \le \frac{\lambda}{|W_1(j\omega)V(j\omega)|}, \quad \omega \in \mathbb{R}, \tag{47}$$

$$|U(j\omega)| \le \frac{\lambda}{|W_2(j\omega)V(j\omega)|}, \quad \omega \in \mathbb{R}. \tag{48}$$

By choosing the functions, W_1, W_2, and V suitably S and U may be made small in appropriate frequency regions.

If the weighting functions are appropriately chosen (in particular, with W_1V large at low frequencies and W_2V large at high frequencies) often the solution of the mixed sensitivity problem has the property that the first term of the criterion dominates at low frequencies and the second at high frequencies:

$$\underbrace{|W_1(j\omega)S(j\omega)V(j\omega)|^2}_{\text{dominates at low frequencies}} + \underbrace{|W_2(j\omega)U(j\omega)V(j\omega)|^2}_{\text{dominates at high frequencies}} = \lambda^2 \tag{49}$$

As a result,

$$|S(j\omega)| \approx \frac{\lambda}{|W_1(j\omega)V(j\omega)|}, \quad \text{for } \omega \text{ small}, \tag{50}$$

$$|U(j\omega)| \approx \frac{\lambda}{|W_2(j\omega)V(j\omega)|}, \quad \text{for } \omega \text{ large}. \tag{51}$$

This allows quite effective control over the shape of the sensitivity and input sensitivity functions, and, hence, over the performance of the feedback system.

Type k control and high-frequency roll-off

In (50)-(51), equality may often be achieved asymptotically. Suppose that $|W_1(j\omega)V(j\omega)|$ behaves as $1/\omega^k$ as $\omega \to 0$. This is the case if $W_1(s)V(s)$ includes a factor s^k in the denominator. Then $|S(j\omega)|$ behaves as ω^k as $\omega \to 0$, which implies a type k control system, with excellent low-frequency disturbance attenuation if $k \ge 1$. If $k = 1$, the system has an integrating action.

Likewise, suppose that $|W_2(j\omega)V(j\omega)|$ behaves as ω^m as $\omega \to \infty$. This is the case if W_2V is nonproper, that is, if the degree of the numerator exceeds that of the denominator (by m). Then $|U(j\omega)|$ behaves as ω^{-m} as $\omega \to \infty$. From $U = -C/(1 + PC)$ it follows that

$C = -U/(1 + UP)$. Hence, if P is strictly proper and $m \geq 0$, also C behaves as ω^{-m}, and $T = PC/(1 + PC)$ behaves as $\omega^{-(m+e)}$, with e the pole excess of P. This means that by choosing m we pre-assign the high-frequency roll-off of the compensator transfer function, and that of the complementary and input sensitivity functions. This is important for robustness against high-frequency unstructured plant perturbations.

Partial pole placement

There is a further important property of the solution of the mixed sensitivity problem that needs to be discussed before considering an example. This concerns a pole cancellation phenomenon that is sometimes misunderstood. First note that the equalizing property implies that

$$\begin{aligned} &W_1(s)W_1(-s)S(s)S(-s)V(s)V(-s) \\ &\quad + W_2(s)W_2(-s)U(s)U(-s)V(s)V(-s) = \lambda^2, \end{aligned} \tag{52}$$

for all s in the complex plane. Next we write the transfer function P and the weighting functions W_1, W_2, and V in rational form as

$$P = \frac{N}{D}, \quad W_1 = \frac{A_1}{B_1}, \quad W_2 = \frac{A_2}{B_2}, \quad V = \frac{M}{E}, \tag{53}$$

with all numerators and denominators polynomials. Note that at this point we do not necessarily take the denominator of V equal to D as before. Then if the compensator transfer function is represented in rational form as $C = Y/X$ it easily follows that

$$S = \frac{DX}{DX + NY}, \quad U = \frac{DY}{DX + NY}. \tag{54}$$

The denominator

$$D_{cl} = DX + NY, \tag{55}$$

is the closed-loop characteristic polynomial of the feedback system. Substituting S and U we easily obtain from (52) that

$$\frac{D^\sim D \cdot M^\sim M \cdot (A_1^\sim A_1 B_2^\sim B_2 X^\sim X + A_2^\sim A_2 B_1^\sim B_1 Y^\sim Y)}{E^\sim E \cdot B_1^\sim B_1 \cdot B_2^\sim B_2 \cdot D_{cl}^\sim D_{cl}} = \lambda^2, \tag{56}$$

where if A is any rational or polynomial function, $A^\sim$ is defined by $A^\sim(s) = A(-s)$.

Since the right-hand side of (56) is a constant, all factors in the numerator of the rational function on the left cancel against corresponding factors in the denominator. In particular, the factor $D^\sim D$ cancels. If there are no cancellations between $D^\sim D$ and $E^\sim E B_1^\sim B_1 B_2^\sim B_2$, the closed-loop characteristic polynomial D_{cl} (which by stability has left-half plane roots only) necessarily has among its roots the roots of D, where any roots of D in the right-half plane are mirrored into the left-half plane.

This means that the open-loop poles (the roots of D), possibly after having been mirrored into the left-half plane, reappear as closed-loop poles. This phenomenon, which is not propitious for a good design, may be avoided, and indeed, turned into an advantage, by choosing the denominator polynomial E of V equal to the plant denominator polynomial D, so that

$$V = \frac{M}{D}. \tag{57}$$

With this special choice of the denominator of V, the polynomial E cancels against D in (56), so that the open-loop poles do not reappear as closed-loop poles.

Further inspection of (56) shows that if there are no cancellations between $M^\sim M$ and $E^\sim E B_1^\sim B_1 B_2^\sim B_2$, and we assume without loss of generality that M has left-half plane roots only, the polynomial M cancels against a corresponding factor in D_{cl}. If we take V proper (which ensures $V(j\omega)$ to be finite at high frequencies) the polynomial M has the same degree as D, and, hence, the same number of roots as D. This means that choosing M is equivalent to reassigning the open-loop poles (the roots of D) to the locations of the roots of M. By suitably choosing the remaining weighting functions W_1 and W_2 these roots may often be arranged to be the dominant poles.

This technique, known as partial pole placement (Kwakernaak, 1986; Postlethwaite *et al.*, 1990), allows further control over the design. It is very useful in designing for a given bandwidth and good time response properties.

In the design examples in Sections 7 and 8 it is illustrated how the ideas of partial pole placement and frequency shaping are combined.

A fuller account of pole–zero cancellation phenomena in $\mathcal{H}_\infty$-optimization problems is given by Sefton and Glover (1990).

Design for robustness

As we have seen, the mixed sensitivity problem is a promising tool for frequency response shaping. By appropriate choices of the functions V, W_1, and W_2, the sensitivity function may be made small at low frequencies and the input sensitivity function (or equivalently, the complementary sensitivity function) small at high frequencies. These are necessary requirements for the system to perform adequately, that is, to attenuate disturbances sufficiently given the plant capacity and presence of measurement noise.

On the other hand, as seen at the end of Section 5, a small sensitivity function S at low frequencies provides robustness against low-frequency perturbations in the plant denominator while a small complementary sensitivity function T at high frequencies protects against high-frequency perturbations in the plant numerator. Investigation of the low-frequency behavior of S and the high-frequency behavior of T permits to estimate the maximal size of the allowable perturbations. Conversely, any information that is available about the size of the perturbations may be used to select the weighting functions V, W_1 and W_2. The choice of these functions generally involves considerations about both performance and robustness. These design targets are not necessarily incompatible or competitive.

It is clear that the crossover region is critical for robustness. The crossover region is the frequency

region where the magnitude of the loop gain $L = PC$—which generally is large for low frequencies and small for high frequencies—crosses over from values greater than 1 to values less than 1. In this region, neither the sensitivity function nor the complementary sensitivity function is small.

Only for minimum-phase plants—that is, plants whose poles and zeros are all in the left-half plane—the sensitivity and complementary sensitivity functions can be molded more or less at will. The presence of right-half plane zeros or poles imposes important constraints. Right-half plane zeros constrain the bandwidth below which effective disturbance attenuation is possible, that is, for which S can be made small. In fact, the largest achievable bandwidth is determined by the right-half plane zero closest to the origin. Right-half plane poles limit the bandwidth of the complementary sensitivity function T, that is, the frequency above which T starts to roll off. Here, the smallest possible bandwidth is determined by the right-half plane pole furthest from the origin. If the plant has both right-half plane poles and right-half plane zeros the difficulties are aggravated. Especially if the right-half plane poles and zeros are close considerable peaking of the sensitivity and complementary sensitivity functions occurs

These results are discussed at length by Engell (1988) for the SISO case and Freudenberg and Looze (1988) for the scalar and multivariable cases. The results show that plants with right-half plane poles and zeros have serious robustness handicaps.

7 EXAMPLE 1: DOUBLE INTEGRATOR

In this section we illustrate the application of the mixed sensitivity problem to a textbook style design example that is simple enough to be completely transparent. Consider a SISO plant with nominal transfer function

$$P_0(s) = \frac{1}{s^2}. \tag{58}$$

The actual, perturbed plant has the transfer function

$$P(s) = \frac{\quad}{s^2(1+s\theta)}, \tag{59}$$

where g is nominally one and the parasitic time constant θ is nominally 0.

We start with a preliminary robustness analysis. The variations in the parasitic time constant θ mainly cause high-frequency perturbations, while the low-frequency perturbations are primarily the effect of the variations in the gain g. Accordingly, we model the effect of the time constant as a numerator perturbation, and the gain variations as denominator perturbations, and write

$$P(s) = \frac{1}{1+s\theta} \tag{60}$$

Correspondingly, the relative perturbations of the denominator and the numerator are

$$\frac{D(s) - D_0(s)}{D_0(s)} = \frac{1}{g} - 1, \tag{61}$$

$$\frac{N(s) - N_0(s)}{N_0(s)} = \frac{-s\theta}{1+s\theta}. \tag{62}$$

The relative perturbation (61) of the denominator is constant over all frequencies, hence also in the crossover region. Because the plant is minimum-phase, trouble-free crossover may be achieved (that is, without undue peaking of the sensitivity and complementary sensitivity functions) and, hence, we expect that—in the absence of other perturbations—variations in $|1/g - 1|$ up to almost 1 will be tolerated.

The size of the relative perturbation (62) of the numerator is less than 1 for frequencies below $1/\theta$, and equal to 1 for high frequencies. To prevent destabilization it is advisable to make the complementary sensitivity small for frequencies greater than $1/\theta$ As the complementary sensitivity starts to decrease at the closed-loop bandwidth, the largest possible value of θ dictates the bandwidth. Assuming that performance requirements specify the system to have a closed-loop bandwidth of 1, we expect that—in the absence of other perturbations—values of the parasitic time constant θ up to 1 will not destabilize the system.

Thus, both for robustness and performance, we aim at a closed-loop bandwidth of 1 with small sensitivity at low frequencies and a sufficiently fast decrease of the complementary sensitivity at high frequencies with a smooth transition in the crossover region. To accomplish this with a mixed sensitivity design, we successively consider the choice of the functions $V = M/D$ (that is, of the polynomial M), W_1 and W_2.

To obtain a good time response corresponding to the bandwidth 1, which does not suffer from sluggishness or excessive overshoot, we assign two dominant poles to the locations $\frac{1}{2}\sqrt{2}(-1 \pm j)$ This is achieved by choosing the polynomial M as

$$\begin{aligned} M(s) &= [s - \tfrac{1}{2}\sqrt{2}(-1+j)][s - \tfrac{1}{2}\sqrt{2}(-1-j)] \\ &= s^2 + s\sqrt{2} + 1, \end{aligned} \tag{63}$$

so that

$$V(s) = s^2 + s\sqrt{2} + 1 \tag{64}$$

We choose the weighting function W_1 equal to 1. Then if the first of the two terms of the mixed sensitivity criterion dominates at low frequencies we have $|S| \approx |\lambda|/|VW_1|$, or

$$|S(j\omega)| \approx |\lambda| \frac{(j\omega)^2}{(j\omega)^2 + j\omega\sqrt{2} + 1} \tag{65}$$

at low frequencies. Figure 11 shows the Bode magnitude plot of the factor $1/V$, which implies a very good low-frequency behavior of the sensitivity function. Owing to the presence of the double open-loop pole at the origin the feedback system is of type 2.

Next contemplate the high-frequency behavior. For high frequencies V is constant and equal to 1.

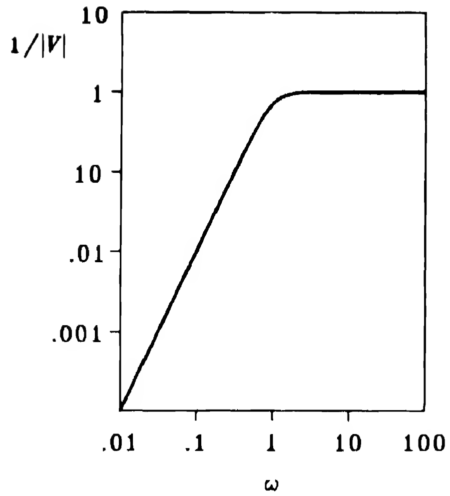


FIG. 11. Bode magnitude plot of $1/V$.

Consider choosing W_2 as

$$W_2(s) = c(1 + rs), \tag{66}$$

with c and r nonnegative constants such that $c \neq 0$. Then for high frequencies the magnitude of W_2 asymptotically behaves as c if $r = 0$, and as $cr\omega$ if $r \neq 0$.

Hence, if $r = 0$, the high-frequency roll-off of the input sensitivity function U and the compensator transfer function C is 0 and that of the complementary sensitivity T is 2 decades/decade (40 dB/decade).

If $r \neq 0$, U and C roll off at 1 decade/decade (20 dB/decade), and T rolls off at 3 decades/decade (60 dB/decade).

We first study the case $r = 0$, which results in a proper but not strictly proper compensator transfer function C, and a high-frequency roll-off of T of 2 decades/decade. Figure 12 shows the sensitivity function S and the complementary sensitivity function T for $c = 1/100$, $c = 1/10$, $c = 1$, and $c = 10$. Inspection shows that as c increases, $|T|$ decreases and $|S|$ increases, which conforms to expectation. The smaller c is, the closer the shape of $|S|$ is to that of Fig. 11.

We choose $c = 1/10$. This makes the sensitivity small with little peaking at the cut-off frequency. The corresponding optimal compensator has the transfer function

$$C(s) = 1.2586 \frac{s + 0.61967}{1 + 0.15563s}, \tag{67}$$

and results in the closed-loop poles $\frac{1}{2}\sqrt{2}(-1 \pm j)$ and -5.0114. The two former poles dominate the latter pole, as planned. The minimal ∞-norm is $\|H\|_\infty = 1.2861$.

Robustness against high-frequency perturbations may be improved by making the complementary sensitivity function T decrease faster at high frequencies. This is accomplished by taking the constant r nonzero. Inspection of W_2 as given by (66) shows that by choosing $r = 1$ the resulting extra roll-off of U, C, and T sets in at the frequency 1. For $r = 1/10$ the break point is shifted to the frequency 10. Figure 13 shows the resulting magnitude plots. For $r = 1/10$ the sensitivity function has little extra peaking while starting at the frequency 10 the complementary sensitivity function rolls off at a rate of 3 decades/decade. The corresponding optimal compensator transfer function is

$$C(s) = 1.2107 \frac{s + 0.5987}{1 + 0.20355s + 0.01267s^2}, \tag{68}$$

which results in the closed-loop poles $\frac{1}{2}\sqrt{2}(-1 \pm j)$ and $-7.3281 \pm j1.8765$. Again the former two poles dominate the latter. The minimal ∞-norm is $\|H\|_\infty = 1.3833$.

Inspection of the two compensators (67) and (68) shows that both basically are PD compensators with high-frequency roll-off. The optimal compensators were computed using a MATLAB package for the solution of $\mathcal{H}_\infty$-optimization problems (Kwakernaak, 1990b) based on the polynomial method of Section 10.

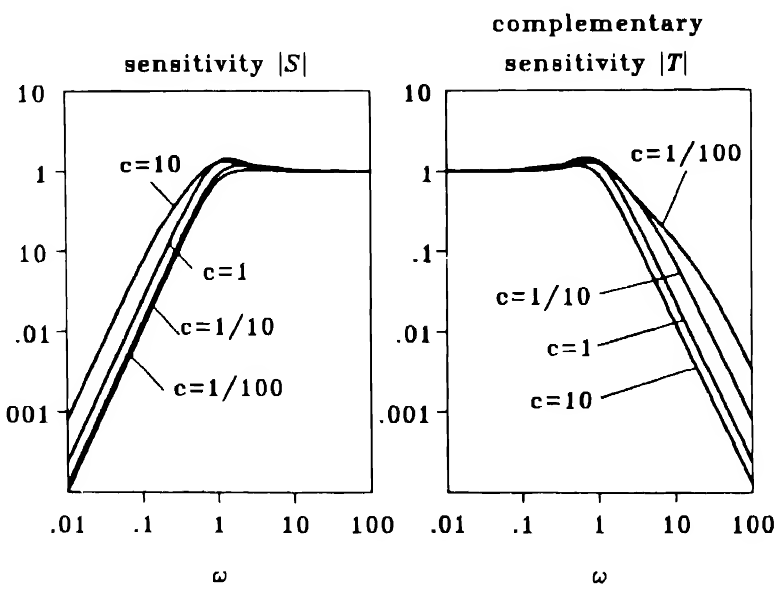


FIG. 12. Bode magnitude plots of S and T for $r = 0$.

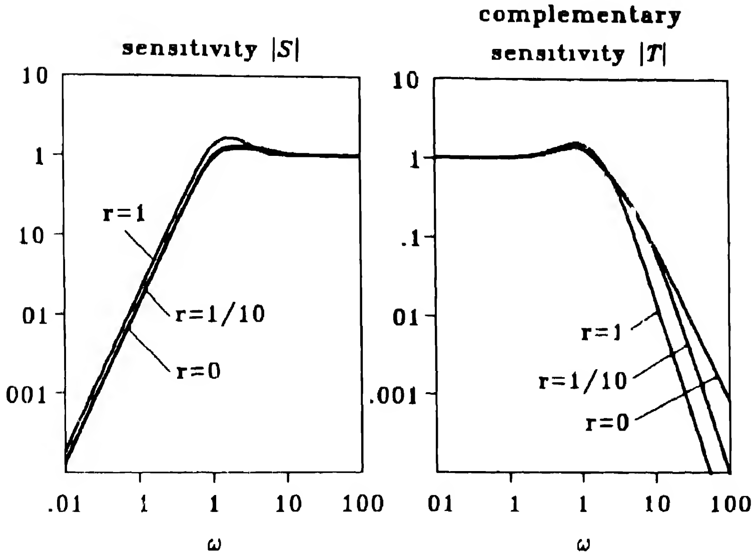


FIG. 13 Bode magnitude plots of S and T for $c = 1/10$

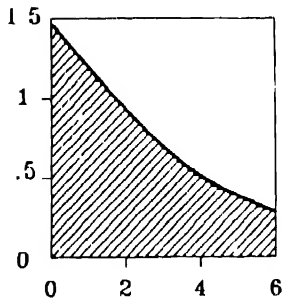


FIG. 14 Stability region

We conclude this example with a brief analysis to check whether our expectations about robustness have come true Given the compensator $C = Y/X$ the closed loop characteristic polynomial of the perturbed plant is $D(s)X(s) + N(s)Y(s) = (1 + s\theta)s^2X(s) + gY(s)$ By straightforward root locus computation,† which involves fixing one of the two parameters g and θ and varying the other, the stability region of Fig 14 may be established for the compensator (67) That for the other compensator is similar The diagram shows that for $\theta = 0$ the closed-loop system is stable for all $g > 0$, that is, for all $-1 < 1/g - 1 < \infty$ This stability interval is larger than predicted For $g = 1$ the system is stable for $0 < \theta < 1.179$, which also is a somewhat larger interval than expected

8 EXAMPLE 2 SHIP COURSE CONTROL

In this section we discuss a more concrete design problem It deals with the heading control of a ship moving at constant velocity, and is included in the 1990 IFAC Benchmark Problems for Control System Design (Åstrom, 1990) Our presentation is related to the discussion in a recent doctoral dissertation by Lundh (1991) The ship transfer function from the rudder angle to the yaw angle is

$$P(s) = \frac{(b_0 s + 1)b_1}{s(s + a_0)(s + a_1)}, \tag{69}$$

TABLE 1 PARAMETER VALUES FOR THE SHIP TRANSFER FUNCTION

Operating conditions	b_0	b_1	a_0	a_1
"Nominal"	0.98	1.72	2.13	0.325
1	1.07	0.75	1.96	−0.70
2	1.05	0.74	1.66	−0.59
3	0.93	0.85	1.86	−0.47
4	0.71	1.29	2.02	−0.21
5	0.89	1.83	2.35	0.05

where the values of the parameters b_0, b_1, a_0, and a_1, depend upon the operating conditions, including speed, trim, and loading In Table 1 the parameter values are given for five operating conditions The table also includes a set of values that were chosen to represent the "nominal" plant The table shows that the sign of the pole a_1 depends on the operating conditions, so that the number of unstable open-loop poles is not constant Lundh (1991) formulates the following design specifications

(1) constant load disturbances at the plant input are rejected at the output,

(2) the closed-loop system is stable at all operating conditions

We first discuss requirement 1, that constant input load disturbances be rejected In the configuration of Fig 1 the transfer function from a load disturbance that is additive to the plant input to the control system output z is

$$R = \frac{P}{1 + PC} = \frac{NX}{DX + NY}, \tag{70}$$

where we write $P = N/D$, $C = Y/X$ Inspection shows that for constant input load rejection the denominator

† At the suggestion of one of the reviewers The author is indebted to this and the other reviewers for many positive and constructive comments

polynomial X of the compensator needs to contain a factor s, that is, the compensator should have integrating action.

Another way of looking at this requirement is to recognize that the transfer function R is related to the sensitivity function S as $R = PS$. For low frequencies the ship transfer function P is proportional to $1/s$. Hence, for R to behave as s, the sensitivity function S should behave as s^2 at low frequencies. In view of the asymptotic analysis of Section 6 this means that the product of weighting functions $W_1 V$ should behave as $1/s^2$. In Section 6 we chose $V = M/D$, $W_1 = A_1/B_1$. For the example at hand, D contains a factor s, so that V already has this factor in the denominator. To provide the additional factor s needed in the denominator of $W_1 V$ we modify the weighting function W_1 to

$$W_1(s) = \frac{\omega_1 + s}{s}\bar{W}_1(s), \tag{71}$$

with ω_1 the frequency up to which the effect of the integrating action extends. $\bar{W}_1$ provides further freedom in selecting W_1. This technique of choosing W_1 to provide integrating action in the compensator also applies to other examples.

Next we discuss the stability robustness specification 2. Since we plan to use the mixed sensitivity design method, we consider which perturbations should be assigned to the plant numerator and which to the denominator. Obviously, the perturbation that causes poles to cross the imaginary axis, that is, the perturbations in the parameter a_1, should be relegated to the denominator. The perturbations in the gain parameter b_1 strongly affect the low-frequency plant characteristics, and therefore are also included in the denominator perturbations. The variations in b_0 principally affect the high-frequency characteristics and therefore are included in the numerator perturbations. The variations in the far-away pole a_0 are mainly important for the high-frequency behavior, but because of the way the pole enters P it also affects the gain. We therefore write the transfer function P in the form

$$P(s) = \frac{b_0 s + 1}{\frac{a_0}{b_1} s\left(\frac{s}{a_0} + 1\right)(s + a_1)} = \frac{b_0 s + 1}{\beta_1 s(\alpha_0 s + 1)(s + a_1)}, \tag{72}$$

where

$$\alpha_0 = \frac{1}{a_0}, \quad \beta_1 = \frac{a_0}{b_1}. \tag{73}$$

Including the variation of α_0 in the numerator perturbations we thus rewrite the plant transfer function as

$$P(s) = \frac{N(s)}{D(s)} = \frac{\frac{b_0 s + 1}{\alpha_0 s + 1}(\bar{\alpha}_0 s + 1)}{\beta_1 (s + a_1)(\bar{\alpha}_0 s + 1)}, \tag{74}$$

where the overbar denotes the nominal value.

It is easy to find that the relative perturbations of

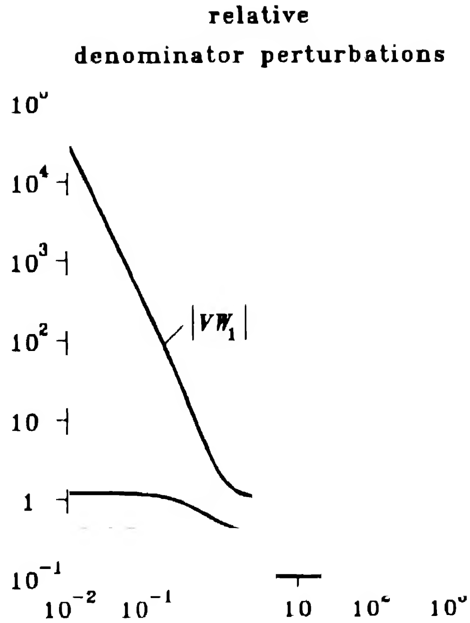


FIG. 15. Magnitudes of relative denominator perturbations for the ship transfer function.

the denominator may be expressed as

$$\frac{D(s) - D_0(s)}{D_0(s)} = \frac{(\beta_1 - \bar{\beta}_1)s + (a_1\beta_1 - a_1\bar{\beta}_1)}{\bar{\beta}_1(s + \bar{a}_1)} \tag{75}$$

with the overbars again denoting nominal values. Figure 15 displays magnitude plots of the relative perturbations for the various parameter combinations of Table 1. We establish a bound for the perturbations in a suitable form. As a preamble to this, we note from Fig. 15 that above the frequency 1 the relative denominator perturbations are less than 1. For this reason, we aim at a closed-loop bandwidth of about 1.

From (37) we see that at low frequencies the relative denominator perturbations need to be bounded by

$$\leq |VW_1| \tag{76}$$

Choosing W_1 as in (71), with $\bar{W}_1(s) = 1$, we consider

$$V(s)W_1(s) = \frac{M(s)}{\bar{\beta}_1 s(s + \bar{a}_1)(\bar{\alpha}_0 s + 1)} \frac{\omega_1 + s}{s} \tag{77}$$

with M a polynomial of degree 3 to be chosen, and ω_1 a constant to be selected. M and ω_1 should be determined such that, first, the bound (76) holds. Second, while doing this we need to remember that the roots of M reappear as closed-loop poles. Third, the constant ω_1 delimits the frequency interval over which the integrating action extends. We choose $\omega_1 = 1$, in line with our decision to choose the closed-loop bandwidth equal to 1. Furthermore, we choose one of the roots of M as $-\bar{a}_0 = -2.13$. This means that the open-loop faraway nominal pole at -2.13 is left in place. Next, we place the two remaining poles of M in a second-order Butterworth configuration with radius 1, that is, we choose this pole pair as $\frac{1}{2}\sqrt{2}(-1 \pm j)$. The choice of the radius is

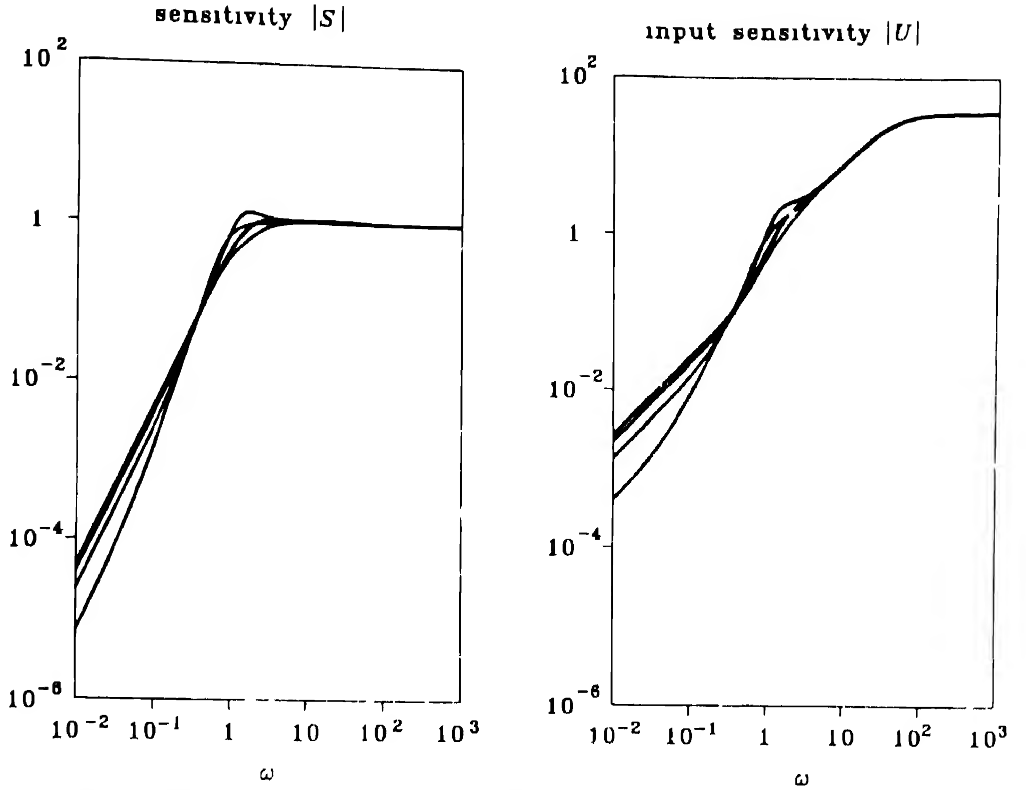


FIG. 16 Magnitudes of the sensitivity and input sensitivity for the ship controller design

again governed by the selection of a closed-loop bandwidth 1. Finally, for normalization we provide M with a leading coefficient such that $V(\infty)W_1(\infty) = 1$.

Thus, we let

$$M(s) = \beta_1(s^2 + \sqrt{2}s + 1)(s - \alpha_0) \tag{78}$$

Figure 15 includes a plot of the magnitude of VW_1. If this magnitude is a bound for the relative denominator perturbations the closed-loop system is robustly stabilized if the minimal value of the mixed sensitivity criterion is less than 1. If the minimal value λ is greater than 1 the bound needs to be rescaled to VW_1/λ. For the plant at hand, which has no right-half plane zeros, and given the normalization $V(\infty)W_1(\infty) = 1$, from experience it may be expected that the minimal value λ is somewhat larger than 1, say, between 1 and 2. Figure 15 shows that a margin of this order of magnitude is available.

The next logical step is to analyse the numerator perturbations. A cursory exploration that is not reproduced here indicates that the numerator perturbation present no great danger to stability robustness.

The final step in the preparation of the mixed sensitivity procedure is to choose the weighting function W_2. Choosing W_2 constant is expected to make the compensator transfer function proper but not strictly proper. In view of the normalization $V(\infty)W_1(\infty) = 1$, we tentatively let $W_2(s) = 0.01$.

Solution of the mixed sensitivity problem results in a minimal ∞-norm of $\lambda = 1.0607$. The optimal compensator transfer function is given by

$$C(s) = \frac{0.5468(s + 2.1277)((s + 0.4331)^2 + 0.4098^2)}{s(s + 47.9401)(s + 1.0204)} \tag{79}$$

C has a pole at 0, as expected, and the transfer function is proper but not strictly proper, as predicted. The nominal closed-loop poles are -45.2010, -2.1277, $-0.7071 \pm j0.7071$, -1.0173 and -1.0030. They include the pre-assigned poles -2.1277 and $-0.7071 \pm j0.7071$. It may be checked by direct computation of the closed-loop poles that the feedback system remains stable with good stability margins, at all operating points. Figure 16 shows plots of the magnitudes of the sensitivity function S and the input sensitivity function U for the various operating points.

For improved robustness against high frequency unstructured uncertainty it is necessary to make the compensator strictly proper to provide roll off of the complementary sensitivity T and the input sensitivity U. To accomplish this we modify the weighting function W_2 to

$$W_2(s) = 0.01\left(1 + \frac{s}{20}\right) \tag{80}$$

The minimum ∞-norm now is 1.1203 while the compensator transfer function takes the form

$$C(s) = \frac{0.5177(s + 2.1277)((s + 0.4262)^2 + 0.4065^2)}{s((s + 28.4134)^2 + (24.5736)^2)(s + 1.0204)} \tag{81}$$

C is strictly proper. The closed-loop pole at -45.2010 is replaced with a pole pair at $-27.0438 \pm j22.9860$, while the remaining closed-loop poles alter not at all or very little. Figure 17 shows plots of the magnitudes of the system functions. Compared with Fig. 16 there is little change in S but U has the desired roll-off.

The results show that the control system specifications may be met with comfortable margins. They may be tightened by including extra specifications, such as

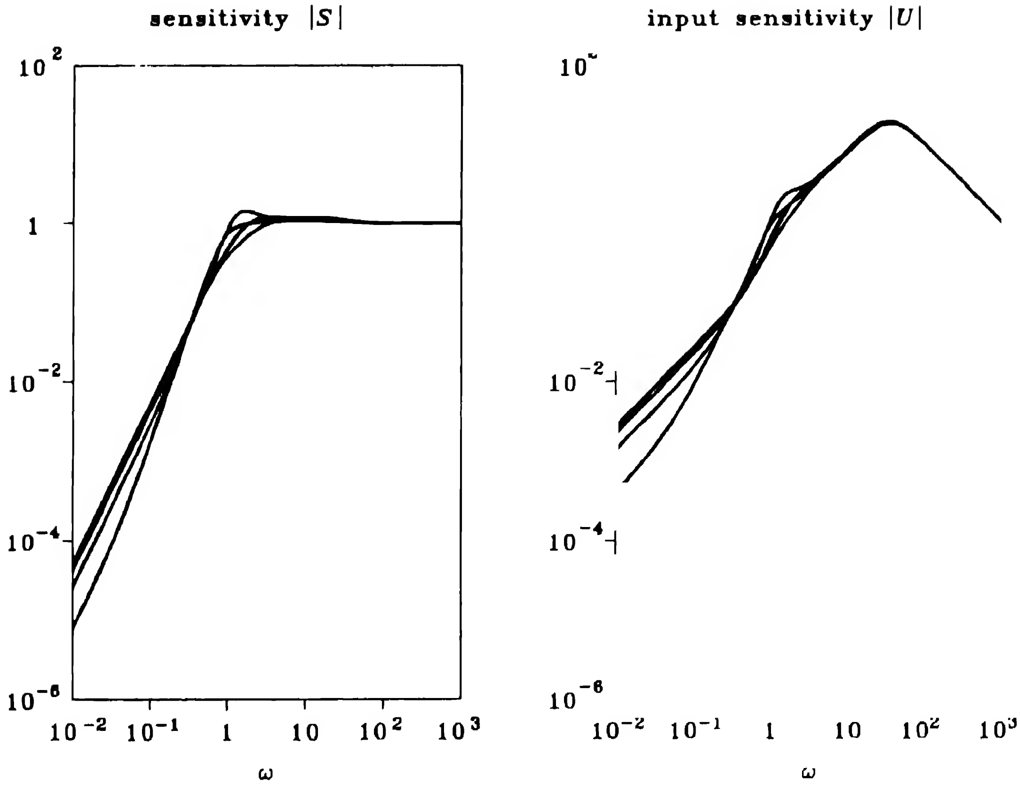


FIG. 17. Magnitudes of the sensitivity and input sensitivity for the second ship controller design.

minimal bandwidth. Lundh (1991) considers additional specifications on the peak values of S and U, on the response to input load disturbances, and on measurement noise sensitivity.

The optimal compensators in this section were again computed using the package based on the polynomial method (Kwakernaak, 1990b), with a small modification needed for the pole at 0 of the weighting function W_1.

9. THE STANDARD $\mathcal{H}_\infty$-OPTIMAL REGULATOR PROBLEM

The mixed sensitivity problem is a special case of the so-called standard $\mathcal{H}_\infty$-optimal regulator problem. We introduce the standard problem by considering the mixed sensitivity function in the configuration of Fig. 18. The diagram shows V as a shaping filter for the disturbance, and W_1 and W_2 as frequency dependent weighting functions for the control system output and the plant input, respectively. The signal w is an external input that drives the disturbance shaping filter V.

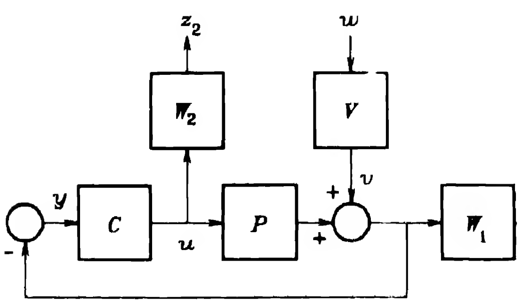


FIG. 18. The mixed sensitivity problem.

It is easy to check that the Laplace transforms of the weighted control system output z_1 and the weighted plant input z_2 are given by

$$\hat{z}_1 = W_1 S V \hat{w},$$
$$\hat{z}_2 = -W_2 U V \hat{w}, \tag{82}$$

so that $\hat{z} = \mathrm{col}\,(\hat{z}_1, \hat{z}_2) = H\hat{w}$, with

$$H = \begin{bmatrix} W_1 S V \\ -W_2 U V \end{bmatrix}. \tag{83}$$

Hence, the mixed sensitivity problem amounts to the minimization of the ∞-norm of the transfer matrix from the external input w to the composite output z. The freedom available in the minimization problem consists of the choice of the compensator C.

By isolating the compensator the block diagram of Fig. 18 may be represented as in Fig. 19, which is the configuration of the "standard" $\mathcal{H}_\infty$-optimal regulator problem. The signal w is an external input, representing driving signals for shaping filters for disturbances, measurement noise, and reference inputs. The signal z represents a control error. Ideally, z is identical to zero. The signal y represents the measured outputs that are available for feedback. The signal u, finally, represent the inputs that may be controlled.

The dynamics of the block G depends on the particular problem at hand. Inspection of Fig. 18 shows that for the mixed sensitivity problem

$$\hat{z}_1 = W_1 V \hat{w} + W_1 P \hat{u},$$
$$\hat{z}_2 = W_2 \hat{u}, \tag{84}$$
$$\hat{y} = -V\hat{w} - P\hat{u}.$$

We rewrite this as

$$\hat{z} = \begin{bmatrix} W_1V & W_1P \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} \hat{w} \\ \hat{u} \end{bmatrix}, \qquad (85)$$

which defines the transfer matrix G as

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \left[\begin{array}{c|c} W_1V & W_1P \\ 0 & W_2 \\ \hline -V & -P \end{array}\right] \qquad (86)$$

There are many other control problems that can be cast as special cases of the standard problem, including problems involving measurement noise and two-degree-of-freedom configurations. Of all these potential $\mathscr{H}_\infty$ applications so far only the mixed sensitivity problem has been investigated in any depth.

The standard problem was first discussed by Francis and Doyle (1987) and is treated at length by Francis (1987).

10. FREQUENCY DOMAIN SOLUTION OF THE STANDARD PROBLEM

The bulk of the research on $\mathscr{H}_\infty$-optimization in the 1980s was devoted to the theoretical and mathematical aspects of the solution of the standard and other problems. Although the theory by no means has reached full maturity, algorithms and software are now becoming available for the solution of the standard problem. The software has not achieved a degree of user-friendliness, however, that affords the user to be totally unfamiliar with the details of the algorithms. For this reason we include in this paper a brief survey of two types of algorithms that are available. The present section is devoted to a frequency domain algorithm. The next section deals with state space algorithms.

Because $\mathscr{H}_\infty$-optimization problems are basically frequency domain oriented, it makes sense to consider frequency domain solutions. The simplest of these solutions relies on what is called J-spectral factorization, a notion that actually is basic for all solution methods. We outline this solution. It is based on Kwakernaak (1990a), which in turn is closely related to work by Green (1989).

It is not difficult to find that the closed-loop transfer matrix H of the configuration of Fig. 19, that is, the transfer matrix from w to z, may be expressed as

$$H = G_{11} + G_{12}(I - KG_{22})^{-1}KG_{21}. \qquad (87)$$

It is characteristic for all approaches to the solution of $\mathscr{H}_\infty$-optimization problems that the problem of minimizing $\|H\|_\infty$ is not tackled directly, but that first the question is studied how to determine suboptimal compensators. Suboptimal compensators are compensators that stabilize the closed-loop system and achieve

$$\|H\|_\infty \leq \lambda, \qquad (88)$$

with λ a given nonnegative number. Optimal compensators follow by finding the smallest value of λ for which such compensators exist.

The inequality $\|H\|_\infty \leq \lambda$ is readily seen to be equivalent to

$$H'(-j\omega)H(j\omega) \leq \lambda^2 I, \quad \text{for all } \omega \in \mathbb{R}, \qquad (89)$$

where the inequality is taken in the sense of definiteness of matrices. We write (89) more compactly as

$$H^{\sim}H \leq \lambda^2 I, \text{ on the imaginary axis}, \qquad (90)$$

where if H is a matrix of rational functions $H^{\sim}$ is defined by $H^{\sim}(s) = H'(-s)$.

For reasons of exposition, consider the special case where $H = P - K$, with P the transfer matrix of a given unstable plant, and K a stable transfer matrix to be determined. It is easily recognized that this is a standard problem with $G_{11} = P$, $G_{12} = -I$, $G_{21} = I$, and $G_{22} = 0$. This problem, which is of more mathematical than practical interest, is known as the Nehari problem (Francis, 1987). Substituting $H = P - K$ into (90), we obtain $P^{\sim}P - P^{\sim}K - K^{\sim}P + K^{\sim}K \leq \lambda^2 I$ on the imaginary axis. This in turn we may rewrite as

$$[I \;\; K^{\sim}] \underbrace{\begin{bmatrix} \lambda^2 I - P^{\sim}P & P^{\sim} \\ P & -I \end{bmatrix}}_{\Pi_\lambda} \begin{bmatrix} I \\ K \end{bmatrix} \geq 0, \qquad (91)$$

on the imaginary axis, which defines the rational matrix Π_λ. We now represent the compensator transfer matrix K in the form

$$K = YX^{-1}, \qquad (92)$$

where Y and X are matrices of stable rational functions. By multiplying (91) on the right by X and on the left by $X^{\sim}$ it follows that the inequality $\|H\|_\infty \leq \lambda$ is equivalent to the inequality

$$[X^{\sim} \;\; Y^{\sim}]\Pi_\lambda \begin{bmatrix} X \\ Y \end{bmatrix} \geq 0, \quad \text{on the imaginary axis}. \qquad (93)$$

This simple derivation applies to the Nehari problem. It may be proved that also for the general case the inequality $\|H\|_\infty \leq \lambda$ is equivalent to (93), with the matrix Π_λ defined as

$$\Pi_\lambda = \begin{bmatrix} 0 & I \\ -G_{12}^{\sim} & G_{22}^{\sim} \end{bmatrix} \begin{bmatrix} \lambda^2 I - G_{11}^{\sim}G_{11} & -G_{11}^{\sim}G_{21} \\ -G_{21}^{\sim}G_{11} & -G_{21}^{\sim}G_{21} \end{bmatrix}^{-1} \times \begin{bmatrix} 0 & -G_{12} \\ I & -G_{22} \end{bmatrix}. \qquad (94)$$

The matrix Π_λ is para-Hermitian, that is, $\Pi_\lambda^{\sim} = \Pi_\lambda$. If

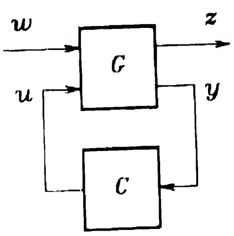


FIG. 19. The standard $\mathscr{H}_\infty$ problem

$\det \Pi_\lambda$ has no poles and zeros on the imaginary axis, Π_λ may be J-spectrally factored as

$$\Pi_\lambda = Z_\lambda^{\sim} J Z_\lambda. \tag{95}$$

Z_λ is a square rational matrix such that both Z_λ and Z_λ^{-1} have all their poles in the open left-half plane. J is a constant matrix of the form

$$J = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \tag{96}$$

with the two I blocks unit matrices of suitable dimensions. J is called the signature matrix of Π_λ.

Given the factorization (95), the condition (93) may be rewritten as

$$[X^{\sim} \quad Y^{\sim}] Z_\lambda^{\sim} J Z_\lambda \begin{bmatrix} X \\ Y \end{bmatrix} \geq 0, \quad \text{on the imaginary axis.} \tag{97}$$

Defining the square stable rational matrix A and the stable rational matrix B by

$$\begin{bmatrix} A \\ B \end{bmatrix} = Z_\lambda \begin{bmatrix} X \\ Y \end{bmatrix}, \tag{98}$$

it follows that (93) is equivalent to

$$A^{\sim} A \geq B^{\sim} B \text{ on the imaginary axis.} \tag{99}$$

By inverting Z_λ we find from (98) that

$$\begin{bmatrix} X \\ Y \end{bmatrix} = Z_\lambda^{-1} \begin{bmatrix} A \\ B \end{bmatrix}. \tag{100}$$

This expression, together with (99), provides an explicit formula for all compensators $K = YX^{-1}$ that make $\|H\|_\lambda \leq \lambda$.

There are many matrices of stable rational functions A and B that satisfy (99). An obvious choice is $A = I$, $B = 0$. This is known as the central solution†.

The question that remains is whether the compensators (100) actually stabilize the closed-loop system. It may be found that a necessary condition for a compensator given by (99)–(100) to stabilize the closed-loop system is that the numerator of $\det A$ have all its roots in the open left-half plane. It may furthermore be proved that if any stabilizing compensator exists that achieves $\|H\|_\infty < \lambda$, all compensators such that $\|H\|_\infty \leq \lambda$ follow from (99)–(100) with A such that $\det A$ has all its roots in the open left-half plane.

These results suggest the following search procedure:

(1) Choose a value of λ.
(2) Determine the J-spectral factor Z_λ and compute a corresponding compensator from (99)–(100) such that $\det A$ has all its zeros in the left-half plane. An obvious possibility is to compute the central solution.
(3) Check if the compensator stabilizes the closed-loop system. If it does, decrease λ. If it does not, increase λ.
(4) If the optimal solution has been approached sufficiently closely, stop. Else, return to (2).

† Note that the central solution as defined here is not unique, because the spectral factorization (95) is not unique.

The rational J-spectral factorization (95) may be reduced to two J-spectral factorizations of polynomial matrices: one for the denominator, one for the numerator. Algorithms for this factorization are now becoming available (Kwakernaak, 1990b; Šebek, 1990; Šebek and Kwakernaak, 1991, 1992).

The search process may terminate in two ways (Kwakernaak, 1990a; Glover *et al.*, 1991). The less common situation is that λ may be decreased steadily until it reaches a lower bound below which the desired J-spectral factorization is no longer possible. All suboptimal compensators for this least possible value of λ then are optimal.

The more usual situation is that λ may be decreased until it reaches a value where the factorization exists but no suboptimal solution stabilizes the closed-loop system. The search procedure may then be used to delimit the optimum. It turns out that as the optimum is approached, the J-spectral factorization becomes singular in the sense that the coefficients of the rational functions occurring in the spectral factor Z_λ grow without bound. At the same time, one of the closed-loop poles of the central solution approaches the boundary of the left-half plane, and actually crosses over from the left-half to the right-half plane, or vice-versa, at the optimal value λ_{opt}. Since the closed-loop transfer matrix H cannot have this closed-loop pole as a pole (because otherwise it would make $\|H\|_\infty$ infinite), this closed-loop pole cancels in H. It turns out that it actually cancels within the compensator transfer matrix C, and hence may be removed.

The singularity phenomenon in the J-spectral factorization may be avoided by only performing a partial factorization, which then may be exploited to compute exactly optimal solutions. The details are described elsewhere (Kwakernaak, 1990a), where also a characterization is given of all optimal solutions, similar to the characterization (100) of all suboptimal solutions. An experimental MATLAB macro package is available for the numerical computation of the optimal solutions (Kwakernaak, 1990b).

There are a large number of details that are not discussed here for lack of space. They concern assumptions on the dimensions of the signals w, u, z, and y, and on the transfer matrix G. Many of these assumptions may be removed or circumvented.

The singularity and cancelation phenomenon does not always occur. If it does not, optimal solutions are obtained corresponding to the largest value of λ such that $\det \Pi_\lambda$ has a pole or zero on the imaginary axis.

The suboptimal and optimal solutions normally are by no means unique. An exception is the SISO mixed sensitivity problem (Kwakernaak, 1990a).

11. STATE SPACE SOLUTION OF THE STANDARD PROBLEM

The mainstream work on algorithms for the solution of the standard problem focuses on state space algorithms (Doyle *et al.*, 1989; Glover and Doyle,

1989). We limit our exposition (which follows that of Weiland, 1990) to a special situation, whose solution admits a neat and compact presentation. The starting point is the description of the system

$$\begin{bmatrix} \hat{z} \\ \hat{y} \end{bmatrix} = G \begin{bmatrix} \hat{w} \\ \hat{u} \end{bmatrix}, \tag{101}$$

in state form as

$$\begin{aligned} \dot{x} &= Ax + Bu + Ew_1, \\ z &= \begin{bmatrix} Dx \\ u \end{bmatrix}, \\ y &= Cx + w_2, \end{aligned} \tag{102}$$

with A, B, C, D, and E constant matrices. This state space representation is not completely general. Note the special structure of the control error z and the fact that the external signal w splits into two separate components $w = \mathrm{col}\,(w_1, w_2)$. For a more general formulation see Glover and Doyle (1989). Note that even in these more general representations the transfer matrix G is limited to be proper, a restriction that is not necessary in the frequency domain solution.

First consider suboptimal $\mathscr{H}_\infty$ regulation using state feedback, that is, when $y = x$. It turns out that in this case $\|H\|_\infty < \lambda$, if at all possible, may be achieved by static linear state feedback of the form

$$u = -Fx, \tag{103}$$

with F a constant matrix. The gain matrix F is given by

$$F = B^T X, \tag{104}$$

where the symmetric matrix X is a nonnegative-definite solution of the algebraic matrix Riccati equation

$$A^T X + XA + D^T D - X\left(BB^T - \frac{1}{\lambda^2} EE^T\right)X = 0, \tag{105}$$

such that the matrix $A - (BB^T - (1/\lambda^2)EE^T)X$ has all its eigenvalues in the open left-half plane. If no such solution X exists, there is no stabilizing state feedback such that $\|H\|_\infty < \lambda$.

One way of proving this result is to note that with u given by (103) the closed-loop transfer matrix from w to z is

$$H(s) = \begin{bmatrix} D(sI - A + BF)^{-1}E \\ -F(sI - A + BF)^{-1}E \end{bmatrix} \tag{106}$$

Manipulation of the Riccati equation (105) in a way similar to the proof of the well-known Kalman–Yacubovitch equality (Kalman, 1964) results in the expression

$$\left[\lambda I - \frac{1}{\lambda} E^T(-sI - \bar{A}^T)^{-1} XE\right]\left[\lambda I - \frac{1}{\lambda} E^T X(sI - \bar{A})^{-1}E\right] = \lambda^2 I - H^T(-s)H(s), \tag{107}$$

where $\bar{A} = A - BF$. If $A - (BB^T - (1/\lambda^2)EE^T)X$ has all its eigenvalues in the open left-half plane, the left-hand side of (107) is positive-definite on the imaginary axis, which proves that $\|H\|_\infty < \lambda$.

Thus, in the case of full state information one algebraic Riccati equation needs to be solved, and static state feedback solves the problem. The output feedback problem, with measurement

$$y = Cx + w_2, \tag{108}$$

is more difficult to solve, although its solution is quite elegant and has a separation structure reminiscent of the LQG problem. It turns out that for output feedback the suboptimal solution needs to be modified to the feedback law

$$u = -F\hat{x}, \tag{109}$$

with $F = B^T X$ as before. The quantity $\hat{x}$ may be viewed as the estimated state, and is the output of an observer-type system given by

$$\dot{\hat{x}} = \left(A - \frac{1}{\lambda^2} EE^T\right)\hat{x} + Bu + ZYC^T(y - C\hat{x}). \tag{110}$$

The symmetric matrix Y, if any exists, is a nonnegative-definite solution of the algebraic Riccati equation

$$AY + YA^T + E^T E - Y\left(C^T C - \frac{1}{\lambda^2} D^T D\right) = 0, \tag{111}$$

such that the matrix $A - Y(C^T C - (1/\lambda^2)D^T D)$ has all its eigenvalues in the open left-half plane. The constant matrix Z in (110) is given by

$$Z = \left(I - \frac{1}{\lambda^2} YX\right)^{-1}. \tag{112}$$

The compensator defined by (109)–(110) is suboptimal and stabilizes the feedback system if and only if $\|XY\|_\infty < \lambda$.

The order of the (suboptimal) compensator equals that of the "plant" G. Representations of "all" suboptimal solutions are also available (Glover and Doyle, 1989). The Riccati equations (105) and (111) are the equivalents of the two polynomial J-spectral factorizations in the frequency domain solution. The Riccati equations are normally solved by spectral decomposition of the corresponding Hamiltonian matrix. Numerically reliable routines are available in MATLAB.

An implementation of a search procedure to delimit the optimal solution analogous to that for the frequency domain approach is available commercially as part of the MATLAB Robust Control Toolbox (Chiang and Safonov, 1988), and the more recent MATLAB μ-Analysis and Synthesis Toolbox. As for the polynomial package, considerable expertise is needed for the use of these toolboxes. As the optimum is approached singularities occur that are similar to those for the frequency domain solution. Glover *et al.* (1991) have analysed these phenomena.

The state space solution of the $\mathscr{H}_\infty$ problem requires more assumptions (for instance that the transfer matrix G be proper) than the frequency domain solution. On the other hand, the numerical algorithms for solving Riccati equations are better developed than the J-spectral factorization algorithms needed in the frequency domain approach.

12. CONCLUSIONS

$\mathcal{H}_\infty$-optimal regulation is a rewarding research subject both for theoreticians and engineers. Theoreticians and also mathematicians find an unequaled opportunity to penetrate deeper into the rich and intricately structured world of linear systems. Engineers recognize many of the issues and design targets of "classical" control theory, which now can be handled algorithmically.

Although the subject has received much attention it has not reached maturity. There are several important topics that we have not been able to touch upon in this tutorial exposition, and are subjects of intensive research.

On the theoretical side, the discrete-time $\mathcal{H}_\infty$-optimal regulation problem is more or less understood. Francis (1990) is doing interesting work on the application of $\mathcal{H}_\infty$-theory to sampled-data systems. Theoretical work on distributed-parameter system is in progress (Curtain, 1991), and attempts are being made to deal with nonlinear problems (Van der Schaft, 1990).

Further theoretical work on finite-dimensional linear systems is directed towards exploring the connections between various solution methods such as those based on the state space approach, J-spectral factorization, operator theoretic methods, interpolation theory, and differential game theory. It cannot be claimed that all aspects of optimal solutions (as opposed to suboptimal solutions) are fully understood, and no doubt considerable attention remains to be spent on this topic.

A problem of considerable interest, where relatively little progress has been made, is how to use the freedom still present in $\mathcal{H}_\infty$-optimal solutions resulting from the lack of uniqueness. The reason for this lack of uniqueness is that the ∞-norm involves the peak value of the largest singular value of the closed-loop frequency response matrix only. This leaves considerable freedom in the behavior of the smaller singular values. The control theoretical interpretation of this freedom is not clear. One way of eliminating the nonuniqueness is to look for solutions among all $\mathcal{H}_\infty$-optimal solutions that successively minimize the peak values of all lesser singular values. This leads to the notion of superoptimality (see e.g. Kwakernaak, 1986; Jaimoukha and Limebeer, 1991).

Another way of eliminating nonuniqueness is to choose so-called "minimum-entropy" solutions (Mustafa and Glover, 1990). Other researchers use the remaining freedom for further optimization purposes.

Another line of research is directed towards making the theory applicable. As we have shown, the SISO mixed sensitivity problem has considerable design potential. The multivariable mixed sensitivity problem shares this, but not all the conclusions for the SISO case generalize straightforwardly. Other special cases of the standard problem, such as criteria involving all three of the sensitivity function, the complementary sensitivity function, and the input sensitivity function, are being looked into. A monograph has been devoted to a special version of the mixed sensitivity problem deriving from what is known as normalized coprime factor plant descriptions (McFarlane and Glover, 1990).

A further question that by no means has been settled is how to translate practical information about plant uncertainty and modeling inaccuracy into quantitative terms that allow the application of $\mathcal{H}_\infty$ techniques. Doyle's "structured singular value" (Doyle, 1982) no doubt is an important step in the right direction.

The fact that algorithms and software become slowly available strongly stimulates work on "real world" applications. More and more interesting design studies are reported, with encouraging results. Several papers presented at a recent meeting in Cambridge attest to this (see for instance Kellett, 1991; Marshfield, 1991; Walker and Postlethwaite, 1991). Other applications are reviewed by Postlethwaite (1991).

The wealth of results on $\mathcal{H}_\infty$-optimization is finding its way into books. Besides the monographs by Mustafa and Glover (1990) and McFarlane and Glover (1990) a book by Morari and Zafiriou (1989) is attracting considerable attention. A recent text by Doyle *et al.* (1991) introduces some of the $\mathcal{H}_\infty$ material at the level of a second course on control.

REFERENCES

Åström, K. J. (1990). Regulation of a ship's heading. In E. J. Davison (Ed.), *Benchmark Problems for Control System Design*. Theory Committee, IFAC.

Chiang, R. Y. and M. G. Safonov (1988). *Robust-Control Toolbox User's Guide*. The Math Works, South Natick, MA.

Curtain, R. F. (1991). State space approaches to $\mathcal{H}_\infty$-control for infinite-dimensional linear systems. *Proc. Conf. Robust Control System Design Using $\mathcal{H}_\infty$ and Related Methods*, 21–22 March, Cambridge, U.K.

Desoer, C. A. and M. Vidyasagar (1975). *Feedback Systems: Input–Output Properties*. Academic Press, New York.

Doyle, J. C. (1979). Robustness of multiloop linear feedback systems. *Proc. 17th IEEE Conf. Decision and Control*, pp. 12–18.

Doyle, J. C. (1982). Analysis of feedback systems with structured uncertainties. *IEE Proc.*, **129**, 242–250.

Doyle, J. C. (1984). *Lecture Notes*, ONR/Honeywell workshop on advances on multivariable control, Minneapolis, MN.

Doyle, J. C., B. A. Francis and A. Tannenbaum (1991). *Feedback Control Theory*. Macmillan, New York.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Trans. Aut. Control*, **34**, 831–847.

Engell, S. (1988). *Optimale Lineare Regelung*. Fachberichte Messen, Steuern, Regeln, **18**. Springer-Verlag, Berlin.

Francis, B. A. (1987). A course in H_∞ control theory. *Lecture Notes in Control and Information Sciences*, **88**. Springer-Verlag, Berlin.

Francis, B. A. (1990). Lectures on H_∞-control and sampled data systems. In E. Mosca and L. Pandolfi (Eds), *H_∞-Control Theory, Como 1990; Lecture Notes in Mathematics*, **1496**. Springer-Verlag, Berlin.

Francis, B. A. and J. C. Doyle (1987). Linear control theory with an H_∞ optimality criterion. *SIAM J. Control Opt.*, **25**, 815–844.

Freudenberg, J. S. and D. J. Looze (1988). Frequency domain properties of scalar and multivariable feedback systems. *Lecture Notes in Control and Information Sciences*, **104**. Springer-Verlag, Berlin.

Glover, K. and J. C. Doyle (1989). A state space approach to H_∞ optimal control. In H. Nijmeijer and J. M.

Schumacher (Eds), *Three Decades of Mathematical System Theory, Lecture Notes in Control and Information Sciences*, **135**. Springer-Verlag, Berlin.

Glover, K., D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally and M. G. Safonov (1991). A characterization of all solutions to the four block general distance problem. *SIAM J. Control and Optimization*, **29**, 283–324.

Green, M. (1989). H_∞ controller synthesis by J-lossless coprime factorization. Department of Electrical Engineering, Imperial College, London. *SIAM J. Control and Optimization*, (to appear).

Jaimoukha, I. M. and D. J. N. Limebeer (1991). State-space algorithm for the solution of the two-block super-optimal distance problem. *Preprints 30th IEEE Conference on Decision and Control*, Brighton, U.K.

Kalman, R. E. (1964). When is a linear control system optimal? *J. Basic Eng. Trans ASME, Ser. D*, **86**, 51–60.

Kellett, M. G. (1991). Continuous scheduling of H^∞ controllers for a MS760 Paris aircraft. *Proc. Conf. Robust Control System Design Using $\mathcal{H}_\infty$ and Related Methods*, 21–22 March, Cambridge, U.K.

Kwakernaak, H. (1983). Robustness optimization of linear feedback systems. *Preprints 22nd IEEE Conference on Decision and Control*, San Antonio, TX.

Kwakernaak, H. (1985). Minimax frequency domain performance and robustness optimization of linear feedback systems *IEEE Trans Aut Control*, **AC-30**, 994–1004.

Kwakernaak, H. (1986). A polynomial approach to minimax frequency domain optimization of multivariable systems. *Int. J. Control*, **44**, 117–156.

Kwakernaak, H. (1990a). The polynomial approach to $\mathcal{H}_\infty$-optimal regulation. In E. Mosca and L. Pandolfi (Eds), H_∞*-Control, Como 1990; Lecture Notes in Mathematics*, **1496**. Springer-Verlag, Heidelberg.

Kwakernaak, H (1990b). MATLAB macros for polynomial $\mathcal{H}_\infty$ control system optimization. Memorandum No. 881, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands.

Lundh, M (1991). Robust adaptive control Doctoral dissertation, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden

Marshfield, W. B. (1991) Submarine periscope depth keeping using an H-infinity controller together with sea noise reduction notch filters *Proc Conf. Robust Control System Design Using $\mathcal{H}_\infty$ and Related Methods*, 21-22 March, Cambridge, U.K.

McFarlane, D. C. and K Glover (1990). Robust controller design using normalized coprime factor plant descriptions, *Lecture Notes on Control and Information Sciences*, **138**. Springer Verlag, Berlin

Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall, Englewood Cliffs, NJ.

Mustafa, D. and K. Glover (1990). Minimum entropy $\mathcal{H}_\infty$ control, *Lecture Notes in Control and Information Sciences*, **146**. Springer-Verlag, Berlin.

Postlethwaite, I. (1991) Robust control of multivariable systems using $\mathcal{H}^\infty$ optimization. *J.A*, **32**, 8–19.

Postlethwaite, I., M.-C. Tsai and D.-W. Gu (1990). Weighting function selection in H^∞ design. *Preprints, 11th IFAC World Congress*, Tallinn, Vol. **5**, pp. 104–109.

Šebek, M. (1990). An algorithm for spectral factorization of polynomial matrices with any signature. Memorandum No. 912, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands.

Šebek, M. and H. Kwakernaak (1991). J-spectral factorization. *Preprints, 30th IEEE Conf. on Decision and Control*, Brighton, 11–13 December, pp. 1278–1283.

Šebek, M. and H. Kwakernaak (1992). Further numerical methods for J-spectral factorization. *Preprints, Second IFAC Workshop on System Identification and Control*, Prague.

Sefton, J. and K. Glover (1990). Pole/zero cancellations in the general $\mathcal{H}_\infty$ problem with reference to a two block design. *Systems and Control Letters*, **14**, 295–306.

Van der Schaft, A. J. (1992). L_2-Gain analysis of nonlinear systems and nonlinear state feedback H_∞ control. *IEEE Trans. Aut Control*, **37**, 770–784.

Verma, M. and E. Jonckheere (1984). L_∞-compensation with mixed sensitivity as a broadband matching problem. *Systems and Control Letters*, **4**, 125–130.

Vidyasagar, M. (1985). *Control Systems Synthesis - A Factorization Approach* MIT Press, Cambridge, MA.

Vidyasagar, M., H. Schneider and B. A. Francis (1982) Algebraic and topological aspects of feedback stabilization. *IEEE Trans. Aut. Control*, **AC-27**, 880–894

Walker, D. and I. Postlethwaite (1991). Discrete time H^∞ control laws for a high performance helicopter *Proc Conf Robust Control System Design Using $\mathcal{H}_\infty$ and Related Methods*, 21–22 March, Cambridge, U.K.

Weiland, S (1990). State space $\mathcal{H}_\infty$ control theory Minicourse on robust control and state space $\mathcal{H}_\infty$-control theory, 1990 Benelux Meeting on Systems and Control, Veldhoven, The Netherlands.

Zames, G (1979). Feedback and optimal sensitivity: Model reference transformations, weighted seminorms, and approximate inverses. *Proc. 17th Allerton Conf.*, pp. 744–752.

Zames, G. (1981). Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses *IEEE Trans. Aut. Control*, **AC-26**, 301–320.

Automatica, Vol 29, No 2, pp 275 283, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Application of EKF Technique to Ship Resistance Measurement*

GENGSHEN LIU†

A new method of "measuring" the resistance of full size ships by applying system identification technique requires only a simple sea trial during a ship's regular voyage but gives the estimated resistance coefficient with good accuracy.

Key Words—Identifiability, system identification, filtering, Kalman filters, marine systems, nonlinear systems, time-domain analysis

Abstract—The measurement of ship resistance is important in naval architecture and marine hydrodynamics research The method of predicting ship resistance from scaled model tests has been used for over 100 years but the results suffer from "scale effect"

This paper presents a new method of "measuring" the resistance of full size ships by using a system identification technique Extended Kalman Filtering is used for estimating intermediate coefficients in the ship's surge motion equation, then the ship resistance coefficient is derived from the identified results This method only requires a simple sea trial during a ship's regular voyage The ship resistance coefficient as well as the wake fraction and the thrust deduction factor were obtained with excellent accuracy The success of this method should impact the design of ship hulls and propellers and also research in marine hydrodynamics

1 INTRODUCTION

REDUCING THE resistance of ships and making their propulsion systems work efficiently has been one of the most important goals for naval architects. However, the complexity of ship bodies makes it difficult to obtain the resistance of ships either by numerical method or by direct measurement. The method of using smaller scaled ship models to do the experiment, and then extending the model results to real ships has been in use since William Froude first proposed this in 1868. But the difference between the model and the full ship still causes serious errors. Efforts to minimize this scale effect have resulted in the current method for estimating ship resistance which consists of a combination of tests on both the model and the ship. The principle rests on the model test results to get the ship resistance curve through extrapolation and to acquire the information of the interaction between the ship hull and the propeller. Based on these results sea trials of the ship are conducted to measure the resistance of the ship by measuring the ship speed and thrust (Comstock, 1967). However, sea trials on real ships in these procedures are very costly, since the ship has to be taken out of service and special instrumentation and qualified personnel have to be placed aboard. (For example, to get the 'EXXON Philadelphia' on one trip between San Francisco and Veldez, Alaska, would cost more than $500,000, which does not count extra-costs for the delay caused by bad weather and other unpredicted reasons.) Furthermore, this method can only be used for newly-constructed ships. Once a ship is put into service, changes in the condition of the ship hull and in the propeller system due to fouling, corrosion addition of appendages, or hull alterations cannot be monitored. The information about ship resistance during service is the key to the determination of the best time to put a ship into dry dock, which is closely related to the operating efficiency of the shipping company.

Our method of applying a system identification technique provides a new way of estimating the ship resistance from simple sea trials. By applying the Extended Kalman Filtering (EKF) technique, not only the ship resistance coefficient C_R but also the full scaled wake fraction w_s and the full scaled thrust deduction factor t_s can be accurately estimated.

In Section 2, the state equation is derived by applying the theory of marine hydrodynamics. In Section 3, the discussion is focused on the application of EKF and the derivation of the ship resistance coefficient. Brief discussion of sea

* Received 11 December 1990, revised 18 June 1991, revised 7 April 1992, received in final form 30 July 1992 The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor R Boel under the direction of Editor H. Kwakernaak.

† McDermott International Inc., P.O. Box 218218, Houston, TX 77218, U.S A.

trial design is presented in Section 4. The results of applying this technique to the tanker 'EXXON Philadelphia' is presented in Section 5, followed by a conclusion in Section 6.

This article is partially based on the research presented in the author's Sc.D. Thesis (Liu, 1988), interested readers can consult this reference for detail.

2. STATE EQUATION

The study of ship resistance requires only the ship surge motion equation as a state equation.

2.1. *Resistance and resistance coefficient of ships*

The resistance of a ship has several components: the frictional resistance caused by the viscosity of water; the eddy resistance caused by the formation of eddies in the boundary layer; and the wave-making resistance caused by the energy lost in the wave made by surface ships. The magnitude of ship resistance R is proportional to the square of the ship surge speed u:

$$R = \tfrac{1}{2}\rho S C_R u^2,$$

where ρ is the water density, and S is the wetted surface area of the ship body (Comstock, 1967).

2.2. *Thrust and thrust coefficients ηs*

Previous results (Liu, 1988) show that the thrust T provided by a screw propeller can be expressed as

$$T = \eta_{p_1}\rho D^2 u^2 + \eta_{p_2}\rho D^3 un + \eta_{p_3}\rho D^4 n^2,$$

where D is the propeller diameter, n is the rotation rate of the propeller shaft, η_{p_1} is the propeller drag–thrust coefficient, η_{p_2} is the propeller induced drag–thrust coefficient, and η_{p_3} is the propeller lift–thrust coefficient.

Introducing the thrust coefficient $K_t = T/\rho D^4 n^2$ then,

$$K_t = \eta_{p_1} J^2 + \eta_{p_2} J + \eta_{p_3},$$

where $J = u/nD$ is called the advance ratio. The K_t–J curve is widely used for describing the characteristics of ship propellers.

2.3. *State equation*

When a ship keeps its straight course, the external forces are the sum of the thrust force and the resistance force. The application of Newton's law, without considering the interaction between the propeller and the ship hull yields,

$$\begin{aligned}(m - X_{\dot u})\dot u &= T - R \\ &= \rho D^2 \eta_{ps_1} u^2 + \rho D^3 \eta_{ps_2} un \\ &\quad + \rho D^4 \eta_{ps_3} n^2 - \tfrac{1}{2}\rho S C_R u^2,\end{aligned}$$

where m is the mass of the ship, $-X_{\dot u}$ is the added-mass of the ship. From now on, η_{ps}s are used to represent the ship propeller thrust coefficients and the model propeller thrust coefficients are represented by η_{pm}s.

However, the interaction between the propeller and the ship hull is not negligible. Therefore the above equation has to be modified.

First, it is noticed that the advancing speed of the propeller u_A, which is the speed of the flow in front of the propeller, is usually lower than u;

$$u_A = (1 - w)u,$$

where w (wake fraction) is less than one.

Second, because of the interaction between propeller and ship hull, the hydrodynamic pressure at the stern of the ship changes. The suction effect changes the flow past the ship hull and usually increases the resistance of the ship, or equivalently, the thrust force is reduced:

$$T = T_0(1 - t),$$

where T is the effective thrust force acting at the ship, T_0 is the thrust force provided by the ship propeller without the interaction between the ship hull and the ship propeller, and t is the thrust deduct factor.

After modification, the ship surge motion equation becomes

$$\dot u = \frac{\rho D^2 \eta_1^* u^2 + \rho D^3 \eta_2^* un + \rho D^4 \eta_3^* n^2}{m - X_{\dot u}} \qquad (1)$$

where,

$$\eta_1^* = \eta_{ps_1}(1 - t)(1 - w)^2 - \frac{S}{2D^2} C_R,$$

$$\eta_2^* = \eta_{ps_2}(1 - t)(1 - w),$$

$$\eta_3^* = \eta_{ps_3}(1 - t).$$

This is the state equation to be used in the system identification procedure which leads to the estimate of the ship resistance coefficient.

It should be noted that strictly speaking t is not a constant. But as the ship speeds up, t approaches a steady value. Therefore, η^*s can be taken as constants. When a ship is speeding up from zero speed, at the beginning stage, its speed is still very low. The value of t is near to zero. Correspondingly, the values of the η_{ps}s in this case are also different, we take them as $\bar\eta$s. Then equation (1) becomes

$$\dot u = \frac{\rho D^2 \bar\eta_1 u^2 + \rho D^3 \bar\eta_2 un + \rho D^4 \bar\eta_3 n^2}{m - X_{\dot u}}. \qquad (2)$$

The difference of equation (1) and equation (2) is beyond the difference of their appearances.

In the latter, since the value of u is very low while the rps value is high, only $\bar{\eta}_3$ plays an important role. Based on this idea, η_{pm_1}, η_{pm_2} and the model test value of C_R can be used in identifying $\bar{\eta}_3$, since their inaccuracy due to the scale effect has little impact on the accuracy of the estimation of $\bar{\eta}_3$. The value of $\bar{\eta}_3$ should be very close to the value of η_{ps_3}, since $\bar{t}$ is very small

Furthermore, there is a special case in ship surge motion–deceleration with propellers wind-milling. In that case, the advanced ratio J_{wm} is a constant, although both u and n changed. The state equation is then simplified to

$$\dot{u} = \frac{\frac{1}{2}\rho S\bar{C}_R u^2}{m - X_{\dot{u}}}, \tag{3}$$

where

$$\bar{C}_R = \frac{(1-t)\left(\eta_{ps_1}(1-w)^2 + \frac{\eta_{ps_2}(1-w)}{J_{wm}} + \frac{\eta_{ps_3}}{J^2_{wm}}\right)D^2}{\frac{1}{2}S} - C_R,$$

which is called the integrated resistance coefficient. If a ship can operate with its propeller wind-milling, this equation can also be used as the state equation.

In the above equations $-X_{\dot{u}}$ is solved through a numerical method, so that $m - X_{\dot{u}}$ is also known (Liu, 1988).

3 EKF AND COEFFICIENT DERIVATION

3.1 *Application of EKF*

For a nonlinear system with the following state model and measurement model:

$$\begin{cases} \dot{\underline{x}} = \underline{f}(\underline{x}, t) + \underline{\varepsilon}(t) & \underline{\varepsilon}(t) \sim N(\underline{0}, Q(t)) \\ \underline{y}_k = \underline{h}_k(\underline{x}(t_k)) + \underline{\nu}_k & k = 1, 2, \ldots, \underline{\nu}_k \sim N(\underline{0}, R_k) \end{cases}$$

where $\underline{x}$ is the state vector, $\underline{f}$ is a nonlinear function of the state. $\underline{\varepsilon}(t)$ is the process noise which is zero mean Gaussian noise with spectral density matrix $Q(t)$. (Denoted by $\underline{\varepsilon}(t) \sim N(\underline{0}, Q(t))$.) $\underline{y}_k$ is the sampled measurement vector at time t_k. $\underline{h}_k$ is a nonlinear function of the state vector at time t_k. $\underline{\nu}_k$ is the measurement noise which is a white random sequence of zero mean Gaussian random variables with associated covariance matrix R_k (denoted by $N(\underline{0}, R_k)$).

The continuous-discrete extended Kalman filter is as follows: (Gelb, 1982)

$$\dot{\hat{\underline{x}}} = \underline{f}(\hat{\underline{x}}, t),$$

$$\dot{P}(t) = F(\hat{\underline{x}}, t)P + PF^T(\hat{\underline{x}}, t) + Q(t),$$

$$\hat{\underline{x}}_{k(+)} = \hat{\underline{x}}_{k(-)} + K_k[\underline{y}_k - \underline{h}_k(\underline{x}(t_k))],$$

$$P_{k(+)} = [I - K_k H_k^T(\hat{\underline{x}}_{k(-)})]P_{k(-)},$$

$$K_k = P_{k(-)}H_k^T(\hat{\underline{x}}_{k(-)}) \times [H_k(\hat{\underline{x}}_{k(-)})P_{k(-)}H_k^T(\hat{\underline{x}}_{k(-)}) + R_k]^{-1},$$

where $\hat{\underline{x}}$ is the expectation of the state. $\hat{\underline{x}}_{k(-)}$ is the expectation of the state just before update at time t_k. $\hat{\underline{x}}_{k(+)}$ is the expectation of the state just after update at time t_k. $P(t)$ is the estimation error covariance matrix defined by

$$P(t) = E[(\hat{\underline{x}}(t) - \underline{x}(t))(\hat{\underline{x}}(t) - \underline{x}(t))^T]$$

$P_{k(-)}$ and $P_{k(+)}$, are respectively the value of $P(t)$ just before and after the update at time t_k. K_k is the gain matrix at time t_k.

$$F(\hat{\underline{x}}, t) = \frac{\partial \underline{f}(\underline{x}, t)}{\partial \underline{x}}\bigg\|_{\underline{x}=\hat{\underline{x}}},$$

$$H_k(\hat{\underline{x}}_{k(-)}) = \frac{\partial \underline{h}_k(\underline{x}(t_k))}{\partial \underline{x}(t_k)}\bigg\|_{\underline{x}=\hat{\underline{x}}_{k(-)}}.$$

There are three cases for the application of EKF:

(1) To identify η^*s:

$$\begin{cases} \dot{\underline{x}} = \begin{pmatrix} \dfrac{\rho D^2\eta_1^* u^2 + \rho D^3\eta_2^* un + \rho D^4\eta_3^* n^2}{m - X_{\dot{u}}} + \xi_u \\ 0 \\ 0 \\ 0 \end{pmatrix}. \\ y = [1 \quad 0 \quad 0 \quad 0]\underline{x} + \zeta_u, \end{cases}$$

where

$$\underline{x} = \begin{pmatrix} u \\ \eta_1^* \\ \eta_2^* \\ \eta_3^* \end{pmatrix},$$

is the augmented state vector, ξ_u is the process noise, and ζ_u is the measurement noise.

(2) To identify $\bar{\eta}_3$:

$$\begin{cases} \dot{\underline{x}} = \begin{pmatrix} \dfrac{\rho D^2\eta_{pm_1}u^2 + \rho D^3\eta_{pm_2}un + \rho D^4\bar{\eta}_3 n^2 - \frac{1}{2}\rho SC_{R_m}u^2}{m - X_{\dot{u}}} + \xi_u \\ 0 \end{pmatrix} \\ y = [1 \quad 0]\underline{x} + \zeta_u, \end{cases}$$

where C_{R_m} is the model test value of C_R and the augmented state vector is

$$\underline{x} = \begin{pmatrix} u \\ \bar{\eta}_3 \end{pmatrix}.$$

(3) To identify $\bar{C}_R$:

$$\begin{cases} \dot{\underline{x}} = \begin{pmatrix} \dfrac{\frac{1}{2}\rho S\bar{C}_R u^2}{m - X_{\dot{u}}} + \xi_u \\ 0 \end{pmatrix} \\ y = [1 \quad 0]\underline{x} + \zeta_u. \end{cases}$$

The augmented state vector is

$$\underline{x} = \begin{pmatrix} u \\ \bar{C}_R \end{pmatrix}.$$

3.2. *Parameter identifiability*

3.2.1. *Identifiability of intermediate coefficients*. The parameters in a state equation are identifiable if they can be uniquely determined by using given information. Clearly, if we put C_R explicitly in the state equation:

$$\dot{u} = \frac{\rho D^2\eta_{ps_1}u^2 + \rho D^3\eta_{ps_2}un + \rho D^4\eta_{ps_3}n^2 - \frac{1}{2}\rho S C_R u^2}{m - X_{\dot{u}}},$$

then both η_{ps_1} and C_R are unidentifiable, because, these two parameters are coupled in the terms of u^2. However, the intermediate parameters in the state equation are identifiable.

For cases (2) and (3), because there is only one parameter to be identified, it follows that both $\bar{\eta}_3$ and $\bar{C}_R$ are identifiable. For case (1), remember that while the ship is accelerated, the rps n is a constant. Therefore, u is the only variable in the state equation. Furthermore, when u is known at any instant, then $\dot{u}$ is also known. And since η_1^*, η_2^* and η_3^* are the coefficients of terms with different orders of u, they are also uniquely determined.

3.2.2. *Feasibility of deriving* C_R *from the identified results*. Since the scale effect between the real ship propeller and the model propeller is considered, η_{ps_1}, η_{ps_2} and η_{ps_3} are also unknown. Furthermore, since the value of t and w derived from the traditional method are not reliable, they are also considered unknown. Therefore, to get a unique solution of C_R, additional equations are introduced.

(1) Although the K_t curves of the model propeller and the ship propeller are different, the zero K_t point of the two curves correspond to the same value of advance ratio. In other words, if the two curves are drawn in the same coordinate system, they will intersect at $K_t = 0$. (see Fig. 1). The reason is that for both the model propeller and the ship propeller, the attack angles of the flow necessary to create zero propulsion force can be taken as the same, though the Reynold's numbers are different for two propellers. And by the same token, to extend K_t curves to negative J values, they should reach their peak values at the same J value. This means that

$$\frac{\eta_{pm_1}}{\eta_{ps_1}} = \frac{\eta_{ps_2}}{\eta_{pm_2}} = \frac{\eta_{ps_3}}{\eta_{pm_3}},$$

which brings in two extra relations.

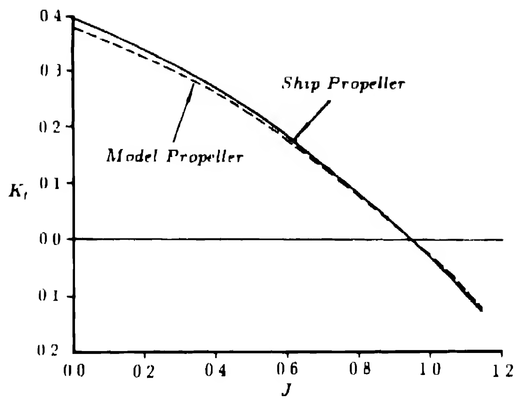


FIG. 1. K_t curves for ship propeller and its model.

(2) According to Abkowitz (1987):

$$t = \frac{\kappa S C_R J_p^2}{D^2 K_t(1-w)}\left(\sqrt{1 + \frac{8K_t}{\pi J_p^2}} - 1\right). \tag{4}$$

where κ depends on the geometry of the ship (hull and propeller).

From this equation, it is clear that t can be calculated at any given value of J_p, if κ is known. We can use either the identified $\bar{\eta}_3$ or the identified $\bar{C}_R$ to calculate κ, since the values of J_p and J_{wm} are both known.

We have just shown that by introducing new equations, the six unknowns η_{ps_1}, η_{ps_2}, η_{ps_3}, w, C_R and κ can be solved for. However, it may be difficult to obtain these unknowns by solving the equations simultaneously. Therefore, two methods of solving for the unknowns using iteration procedures are presented in the following section.

3.3. *Derivation of the resistance coefficient*

As mentioned before, C_R, η_{ps_1}, η_{ps_2}, η_{ps_3}, w and t are combined into the intermediate coefficients η_1^*, η_2^*, η_3^* and $\bar{C}_R$ which are identified by the EKF. The following are two methods for the derivation of C_R, η_{ps_1}, η_{ps_2}, η_{ps_3}, w and t.

3.3.1. *Direct comparison method*. This method is designed to get the ship resistance coefficient C_R through direct comparison between identified values of η^*s, η_{pm}s. During the derivation, t and η_{ps}s can also be determined by using the derived C_R, and w.

(1) First w is determined:

$$w = 1 - \frac{\eta_2^*\eta_{ps_3}}{\eta_3^*\eta_{ps_2}} = 1 - \frac{\eta_2^*\eta_{pm_3}}{\eta_3^*\eta_{pm_2}}. \tag{5}$$

(2) Then η_{t1} is obtained:

$$\eta_{t1} = \eta_{ps_1}\frac{\eta_3^*}{\eta_{ps_3}}\left(\frac{\eta_2^*\eta_{pm_3}}{\eta_3^*\eta_{pm_2}}\right)^2 = \frac{\eta_{pm_1}\eta_2^{*2}\eta_{pm_3}}{\eta_3^*\eta_{pm_2}^2}. \tag{6}$$

(3) The ship resistance coefficient is then determined:

$$C_R = \frac{\eta_{t1} - \eta_1^*}{\frac{S}{2D^2}}. \qquad (7)$$

The determination of C_R and w is completed. With the derived C_R and w, t and η_{ps}s can be derived through iteration.

(4) Derivation of t and η_{ps}s.

Take η_{pm_1} as the first approximation of η_{ps_1}. t_e—the value of t at equilibrium is then determined:

$$t_e = 1 - \frac{\eta_1^*}{\eta_{ps_1}}. \qquad (8)$$

Since the calculation of t_e uses η_{pm_1} instead of η_{ps_1}, it is not accurate and cannot be used to calculate the ship propeller coefficients. In the following, we shall discuss how to derive the values of η_{ps}s while correcting value of t_e through iteration.

Recall from equation (4), that κ can be calculated through

$$\kappa = \frac{t_e D^2 K_{t_e}(1 - w)}{S C_R J_e^2\left(\sqrt{1 + \frac{8K_{t_e}}{\pi J_e^2}} - 1\right)}, \qquad (9)$$

where t_e, J_e and K_{t_e}, are respectively the values of t, J_p and K_t at equilibrium.

During the first approach, t_e is obtained from the model propeller K_t curve instead of the ship propeller K_t curve. This causes the inaccuracy in the calculation of κ, which leads to the inconsistency of the $\bar{t}$ value corresponding to $\bar{\eta}_1$ when it is obtained in the following two ways:

- by comparing $\bar{\eta}_1$ with the model test η_{pm_1}

$$\bar{t} = 1 - \frac{\bar{\eta}_1}{\eta_{pm_1}}, \qquad (10)$$

- by using the formula just established:

$$\bar{t} = \frac{\kappa S C_R \bar{J}_p^2}{D^2 \bar{K}_t(1 - w)}\left(\sqrt{1 + \frac{8\bar{K}_t}{\pi \bar{J}_p^2}} - 1\right). \qquad (11)$$

In order to eliminate this inconsistency, the following steps are needed:

(a) Obtain $\bar{t}$ by substituting $\bar{J}_p$ and $\bar{K}_t$ into equation (11).

(b) Substitute this $\bar{t}$ value into equation (10) to obtain an updated η_{ps_1}.

(c) Substitute η_{ps_1} into equations (5), to get an updated t_e. Then η_{ps_1} and η_{ps_2} are updated by using this updated t_e.

(d) If the updated t_e is unchanged, then stop. Otherwise, repeat the whole procedure until the inconsistency disappears.

This iteration process will lead to a proper t_e value, and at the same time gives a K_t curve for the ship propeller.

3.3.2. *Recurrence method.* This method combines the wind-milling deceleration procedure and the acceleration procedure. The basic idea is comprised of the following steps.

(1) Obtain the advanced ratio of the ship propeller by using the measured data u_{wm} and n_{wm}.

$$J_{wm} = \frac{u_{wm}(1 - w)}{n_{wm} D},$$

where w is obtained from equation (6) by using the acceleration data.

(2) Use η_1^* and η_{pm_1} to calculate t_e.

(3) Recall equation (9), κ can be determined when t_e, J_e and K_{t_e} are known. Once κ is determined, t_{wm} is obtained.

(4) The ship resistance coefficient C_R is then obtained:

$$C_R = \bar{C}_R + \frac{2D^2 K_{t_{wm}}(1 - t_{wm})(1 - w)^2}{J_{wm}^2 S}. \qquad (12)$$

The C_R value is adjusted through the updating of t_e. This is similar to the procedure mentioned earlier.

(a) First calculate $\bar{t}$ based on the C_R value.

(b) Then with the updated $\bar{t}$ value, update t_e, η_{ps_1}, η_{ps_2}, $\bar{K}_t$, $K_{t_{wm}}$, K_{t_e} and κ.

(c) Update C_R.

This procedure is repeated until C_R converges.

4 SEA TRIAL DESIGN AND MEASURED DATA

4.1. *Sea trial design*

Sea trials used for system identification are important, since the accuracy of the identified parameters depends both on the identification technique and on the quality of the data obtained in the experiment.

The sea trial experiment is conducted to provide data with the maximum information content for the system identification under given constraints. The experiment design is an integrated procedure which includes choosing input variables, designing test signals, determining sampling rate and experiment time length, and designing pre-sampling filters.

For the identification of the ship resistance coefficient, the system equation has already been established with n as the input variable. The experiment design is reduced to the choice of the signal pattern for n, the determination of the sampling rate and of the sea trial time length.

4.1.1. *Sea trial pattern.* To get accurate values of the parameters through identification, the

input signal n should have two characteristics. The first characteristic is that for the coefficients η^*s or $\bar{\eta}_3$ or C_R in the model, the input signal n should be able to discriminate between different parameter value groups. In other words, if two different groups of values are assigned to the coefficients, the same input signal n will produce a different output u. The second characteristic is that the variance of the estimates of the parameters should be minimized. The input signals possessing these characteristics are called the persistent exciting signals, and an experiment conducted with that class of input signals is called an informative experiment (Ljung, 1987).

In practice, a perfectly designed experiment plan based on the theoretical analyses may not be feasible due to constraints. Therefore, *a priori* knowledge of the system plays an important role. For the identification of the ship resistance coefficient, the constraint for n is that its value should not be changed too rapidly for the sake of the safety of the propulsion system.

Based on these ideas, two simple maneuvers are considered. The first one is the acceleration maneuver which is to start the propeller from zero rps to full rps. Through this procedure, the measurements of u and n can lead to the estimates of η_1^*, η_2^* and η_3^* simultaneously, and the data collected at the beginning of the acceleration can be used to identify $\bar{\eta}$. The second maneuver is the deceleration or "wind-milling", in which a ship is slowing down with its propeller in the wind-milling mode. However, some ships cannot make the propeller wind-mill, hence for them this maneuver is not applicable.

Checking the two experimental procedures: "speeding-up" and "wind-milling" with simulated data shows that signals of n in both procedures are persistently exciting for the identification of ηs or $\bar{C}_R$.

4.1.2. *Sampling rate and experiment length.* Sampling rate is also related to the accuracy of the identification. For system identification one has to avoid the aliasing phenomenon which occurs when the sampling interval is larger than the time constant of the system (Gustavsson, 1975). The time constant concept is used in linear system analysis. Although, the ship surge motion system is a nonlinear system, for a step input n, the response of u is an exponential-like function. Following the definition of the time constant of a linear system, the pseudo-time constant τ for ship surge motion system can be defined as

$$\tau = \frac{u_e}{\dot{u}_0},$$

where u_e is the value of the u at the equilibrium state, and $\dot{u}_0$ is the value of $\dot{u}$ at time $t = 0$. For the tanker 'EXXON Philadelphia', a pseudo-time constant τ of ~410 sec was obtained both by solving the equation analytically, and by simulation of the ship motion (Liu, 1988). The sample interval of the measurement in the sea trial is 1 sec.

The duration of the experiment should be long enough to insure the reliability of the data and the accuracy of the identification. Insufficient time for an experiment may lead to a pitfall: in some instances, an "accurate" model that fits the data very well may be obtained through identification, however, the model may be wrong. Data from another similar experiment may also lead to an "accurate" model but with quite different parameters obtained through the same identification procedure by the same identification method. Furthermore, the variance of the estimates is usually proportional to the inverse of the experiment length (Gustavsson, 1975). Nevertheless, longer duration may not be better, because extra disturbances are often introduced in a long time experiment, especially during a sea trial. Zarrop (1974), in his thesis, cited an example where time duration is only twice the time constant in practice.

The sea trial time length for each procedure was over 1000 sec. Identification on simulated data showed that this duration is good enough.

4.2. *Measurements*

The sea trial for measuring u and n was conducted on 12 May 1987. The 233 m long, 76,000 dead-weight ton tanker 'EXXON Philadelphia' was on its way back from Valdez, Alaska to San Francisco along the West Coast. The initial position of the sea trial was chosen in calm sea at 58°38.67′ N, and 143°29.25′ W. There was a slight swell of 0.3–0.6 m and no wind.

First, the deceleration procedure was conducted by cutting the steam to the engine and letting the propeller wind-mill. The ship kept moving straight forward. When the ship speed was down to half of the cruising speed, the engine was set in reverse which brought the tanker speed to zero. The acceleration was then initiated by bringing the propeller rotating rate n to 70 rpm as quickly as possible. The whole experiment took about 35 mins and went very smoothly.

5 RESULTS AND DISCUSSION

5.1. *Results of system identification*

During the identification, the basic parameters of the system were set as follows:

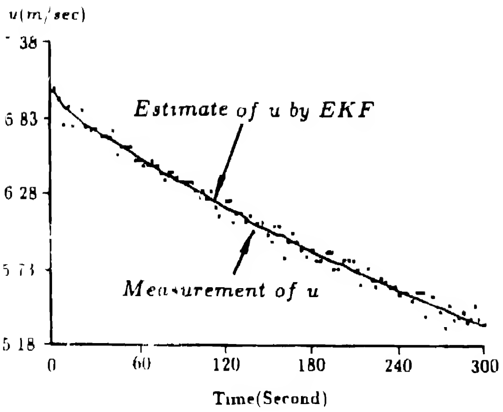


FIG. 2 Estimated u by EKF during the deceleration procedure

- displacement of the tanker(m): 93 000 ton
- added-mass($-X_{\dot{u}}$): 4390 ton
- overall length(L): 233 m
- wetted surface area(S). 1.25×10^4 m^2
- diameter of the ship propeller(D): 8.2 m
- model drag–thrust coefficient(η_{pm_1}): -0.1725
- model associated drag–thrust coefficient (η_{pm_2}): -0.2415
- model lift-thrust coefficient(η_{pm_3}): 0.3796
- model wake fraction(w): 0.383
- model thrust deduction factor(t): 0.216
- water density(ρ): 1.028 ton m^{-3}.

5.1.1. *Result of estimated* $\bar{C}_R$. Two figures are presented here to illustrate the identification procedure and its result. In Fig. 2, the filtering procedure of the EKF is displayed along with the measurement of u. Figure 3 plots the estimated $\bar{C}_R$ in the early path and in the final path of the identification. In the first path, the initial value of $\bar{C}_R$ is a rough guess; while in later paths, the initial value is based on the identification results before it. The validity of the identified results is checked through statistical hypothesis testing (Schweppe, 1976; Hwang, 1980).

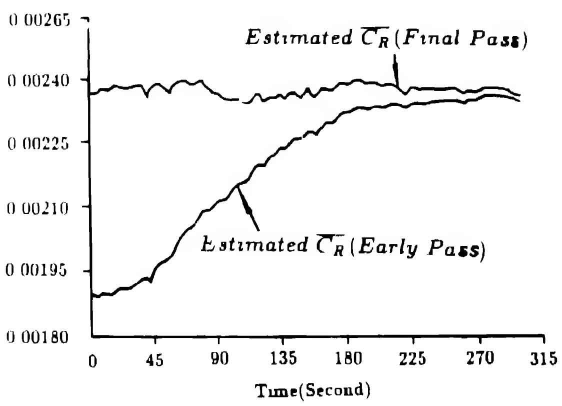


FIG. 3 $\bar{C}_R$ estimated in two paths.

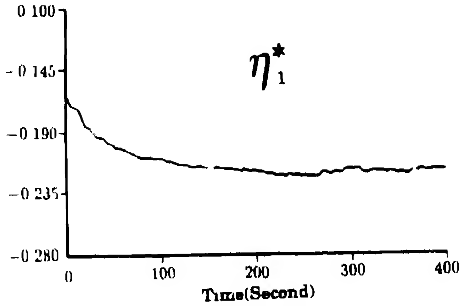


FIG. 4 Estimated coefficient η_1^*

The result of the identification is

$$\bar{C}_R = 0.00238.$$

The advanced ratio J_{wm} is:

$$J_{wm} = 1.03.$$

5.1.2. *Identified* $\bar{\eta}_3$ *at low* J *value*. In the procedure for identifying $\bar{\eta}_3$, the initial part of the measured data was omitted because of the low signal-to-noise ratio. At the same time, to avoid the influence of the increasing thrust deduction factor t during the estimation of $\bar{\eta}_3$, only the data ranging over 200 sec are used at slow speed. Since the input data set is small, the estimation should be conducted carefully. The identified value of $\bar{\eta}_3$ is

$$\bar{\eta}_3 = 0.359$$

5.1.3. *Results of identified* η_1^*, η_2^* *and* η_3^*. Figures 4–6 are plots of the curves of the estimated η_1^*, η_2^* and η_3^* obtained by the EKF technique. The results of the estimation are:

$$\begin{cases} \eta_1^* = -0.285 \\ \eta_2^* = -0.135 \\ \eta_3^* - 0.279. \end{cases}$$

5.2. *Derived coefficient values*

5.2.1. C_R *by direct comparison method*.

(1) The derived values of the hydrodynamic coefficients are:

$$w = 0.239,$$
$$C_R = 0.00227.$$

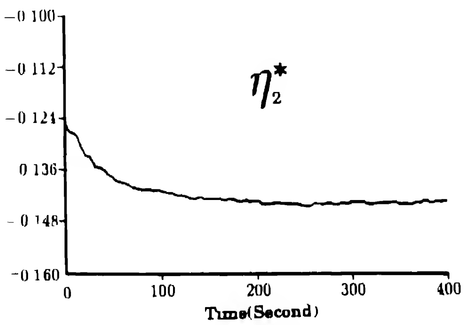


FIG. 5 Estimated coefficient η_2^*

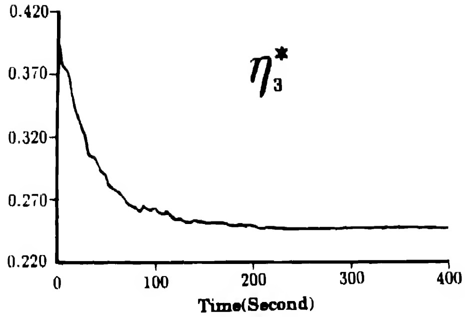


FIG. 6. Estimated coefficient η_3^*.

(2) The first iteration of adjustment of t_e and K_t is as follows:

$$t_e = 0.265,$$
$$K_{t_e} = 0.186,$$
$$\kappa = 0.4179,$$
$$\bar{K}_t = 0.332,$$
$$\bar{t} = 0.085,$$
$$\eta_{ps_3} = 0.3923,$$
$$t_e = 0.2828,$$
$$\eta_{ps_2} = -0.178,$$
$$\eta_{ps_2} = -0.249.$$

Since the value of t_e changed, another iteration is needed.

(3) After the third iteration, t_e converged to a constant. The final results are

$$t_e = 0.296, \quad \eta_{ps_1} = -0.180,$$
$$\eta_{ps_2} = -0.252, \quad \eta_{ps_3} = 0.396.$$

5.2.2. C$_R$ *by the recurrence method.*

(1) From $\bar{C}_R = 0.00238$ and $J_{wm} = 1.03$, the ship propeller thrust coefficient of wind-milling state $K_{t_{wm}}$ is obtained:

$$K_{t_{wm}} = -0.058.$$

(2) Then κ is determined.

$$\kappa = 0.4179.$$

(3) The ship resistance coefficient C_R then is

$$C_R = 0.00214.$$

Through adjusting the ship propeller K_t curve, the updated C_R value is obtained as follows.

(1) Using the value of C_R just obtained to calculate $\bar{t}$:

$$\bar{t} = 0.087.$$

(2) Based on the new value of $\bar{t}$, a new set of values of t_e, η_{ps_1}, η_{ps_2}, $\bar{K}_t$, $K_{t_{wm}}$, K_{t_e} and κ are obtained:

$$\eta_{ps_3} = 0.3932, \qquad t_e = 0.2901,$$
$$\eta_{ps_1} = -0.2077, \quad \eta_{ps_2} = -0.2497,$$
$$\bar{K}_t = 0.3441, \qquad K_{t_{wm}} = -0.048,$$
$$K_{t_e} = 0.196, \qquad \kappa = 0.4538.$$

(3) Then C_R is updated:

$$C_R = 0.00219.$$

The updating procedure is repeated until the C_R value converges to a constant.

For the tanker 'EXXON Philadelphia', the final value of C_R after four iterations by the recurrence method is

$$C_R = 0.00221.$$

When this value is compared with the result obtained from the direct comparison method, the difference is less than 3% of 0.00227. Furthermore, considering the fact that when the ship propeller is wind-milling, the wake fraction w_{wm} is different from that in the acceleration procedure, therefore the C_R value can be further improved through the adjustment of w. For example, if the adjustment is made by referring to the curves given by Huang and Groves (1980), the final result of C_R will be 0.00226. This value will make the difference of C_R between the two methods less than 1% of 0.00227.

To check the results of the estimation, the surge motion of the tanker 'EXXON Philadelphia' is simulated with the estimated hydrodynamic coefficients, and the simulated results are compared with the measured data (see Figs 7 and 8).

The main purpose of this paper is to introduce a new method of measuring ship's resistance. A comprehensive methodology with details of technique for the accurate estimate of resistance

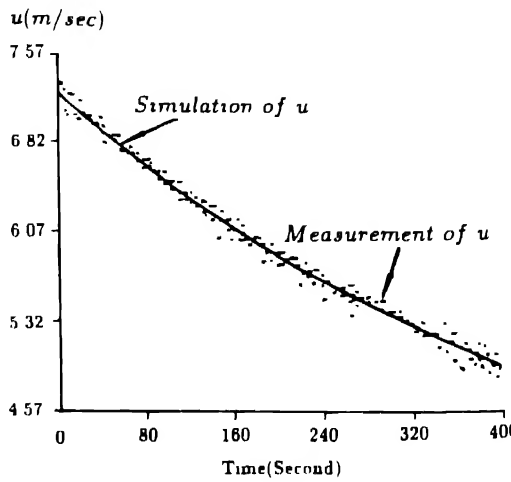


FIG. 7 Simulated u during the deceleration procedure.

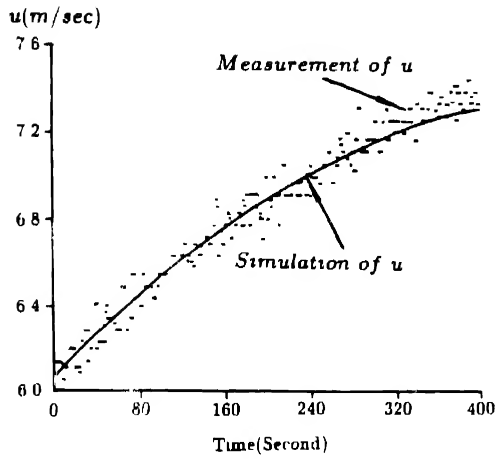


FIG 8 Simulated u during the acceleration procedure

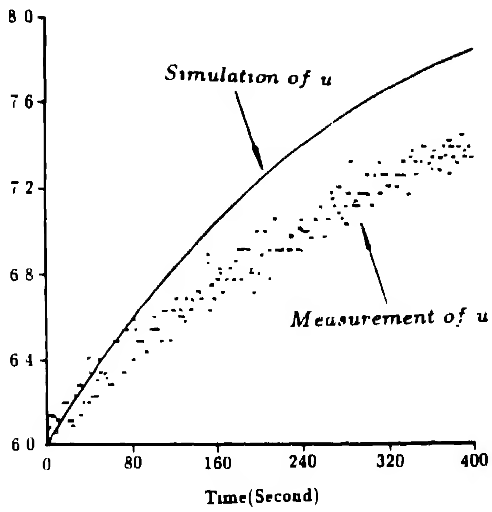


FIG 9 Simulated u with parameter estimate by traditional method

of different kind of ships needs further development. However, the comparison between the simulation using the hydrodynamic coefficients identified by the new method and the traditional method can still show the advantages of the new methods. In Fig 9, the simulation of the surge speed of the ship is compared with the measurement, using $t = 0.221$, $w = 0.38$ and $C_R = 0.25$. The differences between the simulation and the measurements are obvious

6 CONCLUSIONS

The resistance coefficient of the tanker 'EXXON Philadelphia' was successfully estimated by applying EKF technique. The wake fraction w, the thrust deduction factor t, and the ship propeller coefficients were also derived from the identification results and from the model propeller coefficients. The results presented in this paper show that the acceleration process of a ship from zero speed to full speed is sufficient for providing the data necessary for the estimation.

The application of this new method to different kinds of ships will require the modification of system models and redesigning the sea trial. However, the principle of this new method is applicable to most ships. This is not only better for measuring ship resistance than what has been used for over 100 years, which benefits the design of ship hulls and ship propellers, but will contribute to marine hydrodynamics research particularly at the high Reynolds' number.

It should be pointed out that this new method cannot completely replace model testing, not only because model test results constitute a reference for system identification, but also because model tests are used in the design stage to get the hydrodynamic coefficients (in spite of the scale effects). However, with the information derived through the EKF technique, great improvement in the extrapolation methods used to predict ship performance from model test data can be expected (Liu, 1988).

REFERENCES

Abkowitz, M A (1987) Thrust deduction factor during acceleration *Proc ITTC*, Vol 2, pp 151-152

Comstock, J P (Ed) (1967) Resistance and propulsion *Principle of Naval Architecture* The Society of Naval Architects and Marine Engineers, New York

Gelb, A (Ed) (1982) Nonlinear estimation. *Applied Optimal Estimation*, 7th ed MIT Press, Cambridge, MA

Gustavasson, I (1975) Survey of applications of identification in chemical and physical processes *Automatica*, **11**, 3-24

Huang, T and N Groves (1980) Effective wake. theory and experiment *Proc Symp Nav Hydrodyn* , Vol 13, pp 651-673

Hwang, W Y (1980) Application of system identification to ship maneuvering Ph D Thesis, MIT

Liu, G (1988) Measurement of ship resistance coefficient from simple trials during a regular voyage Sc D Thesis, MIT

Ljung, L (1987) *System Identification Theory for the User* Prentice-Hall, Englewood Cliffs, NJ

Schweppe, F C (1976) *Uncertain Dynamic System* Prentice-Hall, Englewood Cliffs, NJ

Zarrop, M B (1974) *Optimal Experiment Design for Dynamics Systems Identification* Springer-Verlag, Berlin

Automatica Vol 29 No 2 pp 285–308 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Grey-box Modelling and Identification Using Physical Knowledge and Bayesian Techniques*

HERBERT J. A. F. TULLEKEN†

Knowledge about plant stability and sign of stationary gains is translated into a set of linear inequalities in the parameters of linear dynamic regression models, producing constrained estimators that are consistent with prior knowledge and attractive under non-ideal experiment conditions.

Key Words—Bayes methods; inequality constrained identification; process industry.

Abstract—Advanced control design requires a model that describes process behaviour adequately. Such a model can be constructed using physical modelling or statistical identification techniques. Both have their disadvantages: physical models are rigorous and thus often expensive to construct, while "black box" model structures are not necessarily compatible with physical reality. Since both approaches have undeniable merits as well, their combination seems to be attractive and rewarding.

This paper discusses an approach to statistical estimation where a "best" linearly parametrized dynamic regression model is identified which is also consistent with specified knowledge about process responses. Such characteristics are forced upon the black box models, yielding a "Grey Box" model set. It will be shown that crucial physical knowledge such as process stability and sign of stationary gains can be translated into linear inequality constraints on the black box model parameters. In order to select "best" estimators in the Grey Box class, a Bayesian approach is adopted. Given a prior distribution associated with the physical knowledge and given the data likelihood, a posterior distribution is constructed. Maximum and average *A Posteriori* estimators are analysed. Explicit solutions are given for special cases of Gaussian likelihood and a prior which is uniformly or piece-wise linearly distributed on a linearly constrained Grey Box model class. Finally, simulation results and an application to a distillation process show the advantage of the contrained estimates under realistic experiment conditions. Considerable variance reductions at the cost of a somewhat larger bias can be achieved, indicating the potential for e.g. adaptive control applications.

1. INTRODUCTION

THE DESIGN of an adaptive or other advanced controller for often multivariable dynamic processes requires a model that gives an accurate description of the process behaviour. Such a model can be constructed by means of rigorous *physical modelling* or using statistical *identification* techniques assuming a "black-box" model structure. Both approaches have their disadvantages: an accurate physical model is hard to get and will be relatively elaborate; the "black-box" model has not necessarily a structure compatible with the underlying physical reality. As a result, conventional (black box model) identification techniques yield unrealistic or even non adequate results in too many cases, e.g. due to modelling errors and non ideal experiments of limited duration. This becomes apparent when characteristics of the model identified contradict common physical knowledge about the process such as stability and sign of static gains. Section 2 discusses these observations, validated with published simulation results. Also an actual application to a distillation column is briefly mentioned to illustrate the physical consistency problems in practice.

A combination of the two approaches might well be attractive and rewarding as a means of combining the obvious advantages. This paper discusses an approach to statistical estimation, aiming at the identification of a linearly parametrized dynamic regression (black-box) model that explains the data "best" under the restriction that its parameter estimates are consistent with prespecified knowledge. Such knowledge is translated into a number of *inequality* restrictions on the black-box model parameters. Together with the regression structure they define a "Grey-Box" class of models that match the given characteristics, which will be discussed and illustrated in Section 3. It will

* Received 3 August 1990; revised 19 June 1991; received in final form 1 May 1992. A condensed version of this paper was presented at the IFAC Symposium on Adaptive Control and Signal Processing which was held in Glasgow, U.K. during April 1989. The Published Proceedings of this IFAC Meeting may be ordered from Pergamon Press Ltd, Headington Hill Hall, Oxford OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor Y. Sunahara under the direction of Editor P. C. Parks.

† Koninklijke/Shell Laboratorium, Amsterdam (Shell Research B.V.), P.O. Box 3003, 1003 AA Amsterdam, The Netherlands.

be shown that crucial features such as (open-loop) stability and bounds on the stationary gains of the process give rise to *linear* inequalities $A\theta \le B$ on the parameter θ. This section relates to earlier work such as that of Jury (1974), Shieh *et al.* (1987), Shieh and Sacheti (1978) and Jørgenson *et al.* (1985). It offers a novel outlook on the translation of practical (plant) knowledge to an admissible parameter set.

In Section 4, a Bayesian approach to Grey-Box parameter estimation is described. Using the Grey-Box model class a prior p.d.f. (probability distribution function) for the model parameters is constructed. This prior is combined with the likelihood using Bayesian techniques yielding a posterior p.d.f. which is conditioned on data as well as on process characteristics.

In order to select a "best" model based on the posterior p.d.f., suitable estimators have to be defined. To this end maximum *A Posteriori* and average *A Posteriori* estimators are discussed in Section 5. These notions are generalizations of standard estimation problems, replacing the likelihood by the Bayesian posterior. Three constrained estimators are highlighted. Moreover, explicit solutions to the estimation problems will be given for the case of Gaussian likelihood, and a prior, which is uniformly or polyhedrally distributed on a linearly restricted Grey-Box model set. Relations exists, with respect to general statistical work, with Akaike (1978, 1986) and Liew (1976), as well as with identification-oriented publications such as in Bard (1974), Gertler (1979), Peterka (1981) and Goodwin and Sin (1984).

In addition, some simulation results will be presented and the techniques will be applied to distillation column data. Finally, in Section 6 conclusions will be drawn.

2. PHYSICAL INCONSISTENCIES OF IDENTIFICATION RESULTS

2.1. *Realistic experimental conditions in the process industry*

In the petrochemical industry the identification of a statistical dynamic process model is often troublesome. This structural problem can arise from violated assumptions for model and estimation structure (e.g. linearity, stationarity) but is often primarily due to insufficient information in the field data. This is the case for instance where, due to safety and/or quality considerations the duration of field experiments and the intensity of test-signal perturbation must be kept to a minimum. Typically, the length of a test-signal perturbed field test will be about three or four times the 99% process settling time whereas the signal-to-noise-ratio will be about unity or slightly better. In the following we will refer to this as *realistic experimental conditions*. As a result popular and theoretically well-understood one-step-ahead prediction error based identification methods such as least squares (LS), maximum likelihood (ML), generalized/extended least squares (ELS), instrumental variables (IV), etc. (see e.g. Eykhoff (1974), Goodwin and Payne (1977), Söderström and Stoica (1989)) often fail to produce physically consistent models from finite, noisy experiments. For realistic experimental conditions but otherwise ideal factors (stationary stochastics; deterministic and noise model structure correctly chosen) Monte-Carlo simulations show that lack of stochastic convergence of these estimators results in a significant percentage of unrealistic or even non-sense models. This does not relate to parameter bias only, but also to the description of the input–output behaviour of the plant (see Wahlberg and Ljung (1986) for frequency-domain and Tulleken (1990) for time-domain characteristics). Since these problems can occur quite easily under rather idealized conditions, one could expect even more trouble in practical situations where model structure uncertainty and non-stationarities come in. This is indeed often the case when realistic experimental conditions apply as the following illustrates.

2.2. *Example* 1: *physical inconsistencies in identified distillation models*

Consider part of a distillation train as shown in Fig. 1. A feed stock consisting of many components is fed halfway into the main column. As a result of the heat injection of the reboiler, components with a high volatility (vapour) escape through the top, are cooled down in the condenser, accumulated in a reservoir and partly fed back into the column for improved separation. On the other hand, components with a low volatility (liquid) flow off to the bottom. From the middle of the column a side flow is withdrawn, with on average, middle-cut components (i.e. with an intermediate boiling point). This material is fed into a smaller "side-stripper", which sends the relatively heavy components into its bottom stream and recycles the lighter components into the main column. A multivariable model was required for controller design, specifying the dynamic relationship between two inputs (i.e. small set point changes on reflux flow $u_1 := \phi_R$ and side-stripper flow $u_2 := \phi_S$) and two outputs, i.e. the qualities

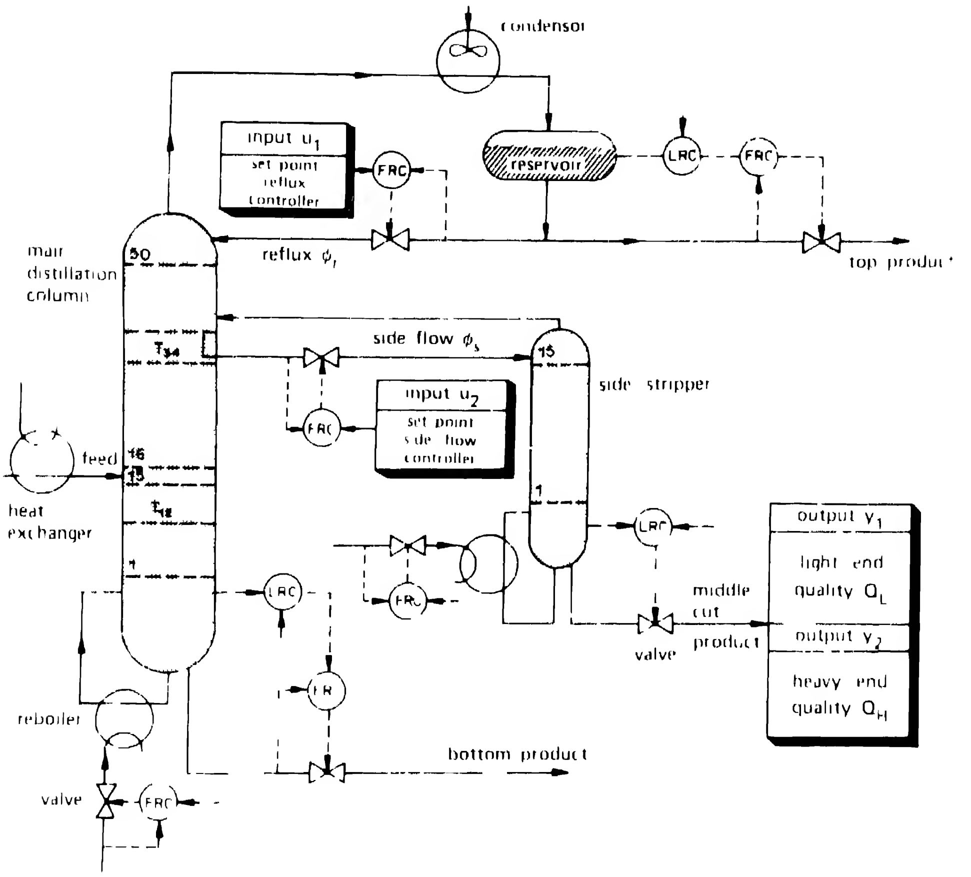


FIG. 1. Relevant fraction of distillation process during data acquisition (FRC – Flow Rate Controller, LRC – Level Rate Controller)

$y_1 := Q_l$ and $y_2 := Q_h$ of lighter and heavier components, both in the bottom flow of the side-stripper. In addition, intermediate outputs were defined: the top temperature T_{34} at tray 34 and the bottom temperature T_{12} at tray 12 of the main column. These easily measurable temperatures have the advantage to be usually highly correlated with the less easily attainable quality measurements. The latter suffer from severe delay (in this case 8 mins) and inaccuracy. Under difficult, but realistic field conditions an experiment was carried out using Pseudo Random binary Noise (PRBN) test signal perturbations on the two inputs (set points), whereas feed rate and composition, set points of reboiler duty and level range controllers were not manipulated. Together with these test-signal realizations, temperature and quality variations were recorded using a relatively short sampling period T_0 of half a minute (the 95% settling time was about 5 h). A 10 h period of zero-averaged, scaled input–output behaviour (1200 samples) is shown in Fig. 2.

As can be seen from the temperature data, the signal-to-noise ratio is unfavourable. This is a consequence of operational constraints imposed on the process excitation. Because of the urge to operate in a safe and economical manner, relatively small excitation and rapid switching of the flow controller set points as well as a short experiment length had to be accepted. Since these conditions are often encountered in process industry, *the challenge is to produce an acceptable model with such limited, noisy experiments*. This argument applies especially to real-time (e.g. adaptive) control applications.

The model class considered for this distillation process consisted of multivariable discrete-time ARMAX(n, m, l, δ) models (Goodwin and Sin (1984), Tulleken (1987)):

$$\begin{aligned} y_k + A_1 y_{k-1} + \cdots + A_n y_{k-n} &= B_1 u_{k-1-\delta} + \cdots + B_m u_{k-m-\delta} \\ &\quad + \xi_k + C_1 \xi_{k-1} + \cdots + C_l \xi_{k-l} + d. \quad (1) \end{aligned}$$

Here y, ξ and u are discrete-time ν-dimensional output, white noise and μ-dimensional input signals, respectively, where $y_k := y(kT)$, etc. and T is the sampling period. Furthermore n, m and l are the dynamic orders while $\delta \in \mathbb{N}$ is the *overall* discrete dead-time index, which is present in all input components. The $\nu \times \nu$ Auto Regressive matrices A, the ν-dimensional Load vector d (modelling a trend component, arising

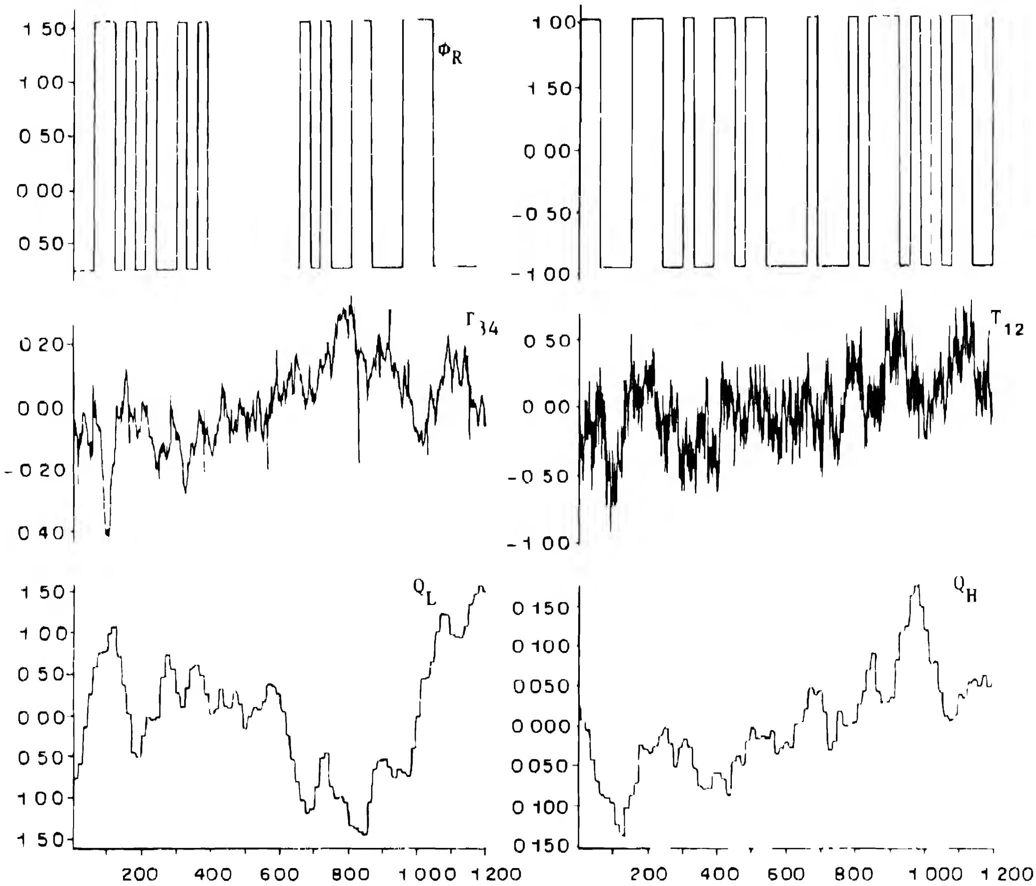


FIG. 2 The experimental flow/temperature/quality data over a period of 10 h

from disturbances, non-zero noise mean and/or linearization of non-linear dynamics) as well as the $\nu \times \nu$ Moving Average matrices C and the $\nu \times \mu$ external matrices B are deterministic (possibly slowly varying), unknown ARMAX parameters.

In this application, $\nu = \mu = 2$ with $u := [\phi_R, \phi_S]^T$ and $y := [T_{34}, T_{12}]^T$, respectively $y := [Q_H, Q_L]^T$. For a number of reasonable choices of the structural parameters n, m, l, δ and T, successive "ELS-best" models of the form (1) have been obtained using the data presented in Fig. 2.

The results were discouraging: only few statistically "most likely" models did represent *essential*, known characteristics of the process at hand, such as stability and correct sign of static gains. A characteristic example for the ELS estimation of the case $n = m = l = 4$ and various choices for overall dead-time δ (resulting in an input–output shift correction for process dead-time) and T (chosen as a multiple κ of the basic sampling time $T_0 = 0.5$ min, by simply ignoring intermediate data) is shown in Fig. 3. Each marked pair (κ, δ) represents an ARMAX(4, 4, 4, δ) ELS-optimal model discretized with sample time κT_0 which, in addition, turned out to be *stable* and representing the *correct signs* of all four static gains (known from physical considerations); all other (κ, δ) pairs corresponded to ELS-optimal models which did not represent all those vital characteristics correctly and were not marked. Two types of models are considered, having either temperatures T_{34} and T_{12} or qualities Q_H and Q_L as outputs. A relatively small number of potentially valuable, provisionally accepted models of each type fall within a "cigar-shaped" region in which non-optimal choices of structural parameters such as T and δ are apparently compensated for, cf. Fig. 3.

However, *almost all* of these models had to be rejected, since additional step-response characteristics (i.e. the knowledge that some loops showed (non)minimum phase/(non)oscillatory dynamics; rough knowledge about settling-times) were not represented correctly. Lower order choices for n, m and l (including ARX only, $l = 0$, using LS identification) yielded quite similar results. Higher order models could not reasonably be considered because of the imbalance between parametrization load (too many parameters) vs the information available (a relatively short and noisy experiment).

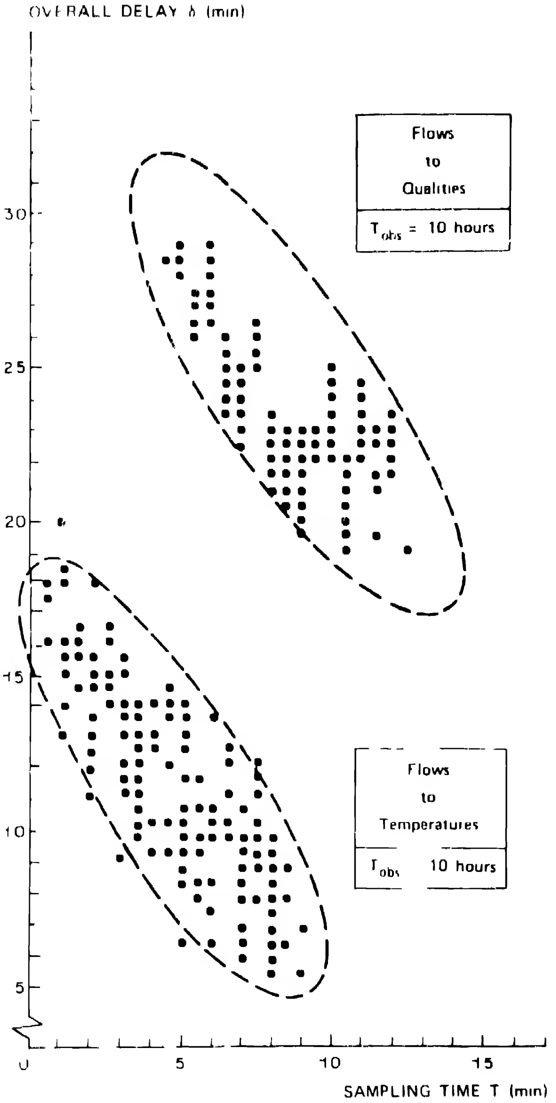


FIG. 3. Dotted cases refer to ELS-identified, with sample time T discretized ARMAX(4, 4, 4, δ) models being stable and having correct sign of all four static gains, using the data shown in Fig. 2. Empty positions refer to ELS-optimal models failing for at least one of those features.

In conclusion, with this relatively noisy and limited data set, *hardly any acceptable* black-box model could be identified.

2.3. *Use of physical knowledge*

It is not our intention to advocate brute force search or discuss the obvious merits of model structure discrimination methods, either statistically (e.g. Akaike, 1974) or using validation with respect to model characteristics. What we would like to stress is that the often heard slogan "let the data speak for themselves" does not apply to these non-ideal (finite, noisy) experiments, even when the process would be in the model class. Due to the lack of information, structural identified model inconsistencies arise, no matter what statistical data analysis method is being used. Still, one has to come up with an acceptable model since performing a prolongued experiment is usually too costly or impossible. In those cases, available physical knowledge about crucial plant characteristics should be supplied and somehow be integrated in the identification method. Moreover, it has been widely accepted that modelling errors are unavoidable but influenceable aspects of identification practice. If at least the most relevant process characteristics are sufficiently accurately reflected, approximate modelling is fully acceptable. In our view this implies the use of short-cut, *low-order* black-box model structures where crucial information is translated into constraints on model parameters, restricting the "black-box" models to a physically more consistent "Grey-Box" model class. Systematic errors (Goodwin and Salgado, 1989) can be circumvented or obstructed with such an approach. A similar line of reasoning applies to the non-linear modelling approach.

3. GREY-BOX MODELLING USING PHYSICAL KNOWLEDGE

3.1. *The Grey-Box model class $\mathcal{M}^{gb}$*

Consider the ARMAX(n, m, l, δ) model class (1), which belongs to a more general black-box class $\mathcal{M}^{bb}$ of *linearly parametrized* multivariable discrete-time regression models

$$\mathcal{M}^{bb} := \{y_k = \theta z_k + \xi_k \mid \theta \in \mathbb{R}^{\nu\times p}; k \in \mathbb{Z}\}, \quad (2)$$

where y_k is the ν-dimensional output vector of the process, θ is a constant, unknown parameter matrix, z_k a known vector of p regressors and ξ_k a ν-dimensional (discrete-time) white noise, all at time k. This can be seen from the following

$$\theta := [-A_1|\cdots|-A_n \mid B_1|\cdots|B_m \mid C_1|\cdots|C_l \mid d],$$

$$z_k^T := [y_{k-1}^T|\cdots|y_{k-n}^T \mid u_{k-1-\delta}^T|\cdots$$
$$\cdots|u_{k-m-\delta}^T \mid \xi_{k-1}^T|\cdots|\xi_{k-l}^T \mid 1].$$

However, in general the regression vector z_k can also be a non-linear function of selected regressors which may be motivated with physical arguments (conservation laws, use of dimensionless groups). Our aim will be to force important physical structure upon the models in $\mathcal{M}^{bb}$ by constraining the candidate model parameters θ to a Grey-Box (GB) model parameter set Ω, defined for suitable (in general matrix-valued) functions f and g as

$$\Omega := \{\theta \in \mathbb{R}^{\nu\times p} \mid f(\theta) = 0; g(\theta) \geq 0\}. \quad (3)$$

This subsequently leads to a *Grey-Box model class*

$$\mathcal{M}^{gb} := \{y_k = \theta_{z_k} + \varepsilon_k \mid k \in \mathbb{Z}; \theta \in \Omega\} \subset \mathcal{M}^{bb}, \quad (4)$$

each model in $\mathcal{M}^{gb}$ being consistent with the *a priori* physical knowledge.

Imposing physical constraints on statistical models is not a new topic. Much has been written for the case of equality constraints ($f(\theta) = 0$), as it often happens that we can fix (a combination of) parameters beforehand (e.g. Goodwin and Payne, 1977). A typical example is in dead-time correction for multivariable ARMAX models where some elements of leading B matrices in (1) are set to zero. A more advanced application uses a Dynamic Quantity Interaction Diagram from which it can easily be seen which parameters in a multivariable ARMAX model must equal zero (see e.g. Jørgensen *et al.*, 1985). However, inequality constraints on black-box model parameters ($g(\theta) \geq 0$) received only little attention in the statistical literature, probably because of the foreseen additional complexity and the "let the data do the job" attitude. For instance, Goodwin and Sin (1984) and Liew (1976) do make some general remarks about estimated parameter consistency. On the other hand, texts more related to physical (non-linear) modelling and estimation (such as Bard, 1974) emphasize the importance of this kind of model structure enhancement. In the following sub-section we will focus on inequality constraints solely. If present the constraint $f(\theta) = 0$ may be reformulated in $[|f(\theta)| - f(\theta)] \geq 0$ which can be included in $g(\theta) \geq 0$. For reasons of simplicity and in view of the estimation problem we restrict ourselves to convex and even *linear* g. Consequently (3) reduces for suitable matrixes A (possibly rank deficient) and B to the convex, possibly unbounded set

$$\Omega := \{\theta \in \qquad | A\theta \leq B\}. \tag{5}$$

Vital, important knowledge such as admissible areas for poles and static gains (often denied by identified model results) can be specified in terms of (5).

3.2. *Inequality constraints from physical knowledge*

In this section we consider the discrete-time transfer function matrix $\mathcal{H}$ of the deterministic part of (1) where z^{-1} is the unit shift backward operator and $\mathcal{A}$ and $\mathcal{B}$ are polynomial matrices:

$$\mathcal{H}(z^{-1}) = [\mathcal{A}(z^{-1})]^{-1}\mathcal{B}(z^{-1}), \tag{6}$$

with $\mathcal{A}(z^{-1}) = I + \sum_{i=1}^{n} A_i z^{-i}$; $\mathcal{B}(z^{-1}) = \sum_{i=1}^{m} B_i z^{-i}$.

We will derive inequality constraints on the parameters of the model (6), necessary for the satisfaction of known properties of the underlying process.

3.2.1. *Admissible pole locations yielding an admissible parameter set.* First we will consider the situation that the plant is known to be open-loop (perhaps including low-level controllers) *stable* in the sense of Lyapunov. In the process industry this holds for the vast majority of the processes, with e.g. exothermic chemical reactors being one of the rare exceptions. In many cases the plant is even known to be asymptotically stable, with additional knowledge about an upper bound for the settling time. We shall aim at the formulation of the sharpest necessary *linear* inequality conditions on the coefficients of $\mathcal{A}$ in the model (6) that emerge from a given admissible area for the poles of

$$\mathcal{A}(z^{-1}) = I + A_1 z^{-1} + A_2 z^{-2} + \cdots + A_n z^{-n}, \tag{7}$$

i.e. the roots of det $\mathcal{A}(z^{-1}) = 0$. In other words, we are interested in the smallest, *linearly bounded* convex hull of the admissible parameter set. In the general *multivariable* case one is rather restricted with respect to deriving necessary linear constraints since the associated admissible parameter set will often be unbounded and non-convex. Since stability is an important special case, the following may serve as a simple illustration.

3.2.2. *Example 2: stability constraints for two-dimensional first-order AR system.* Consider a first-order discrete-time AR model with two outputs described by:

$$y_k + Ay_{k-1} = y_k + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} y_{k-1} = 0.$$

Its asymptotic stability is equivalent with the roots of det $\mathcal{A}(z) = \det(zI + A) = z^2 + a_1 z + a_2 = 0$ being strictly within the unit circle, where $a_1 :=$ trace $A = a_{11} + a_{22}$ and $a_2 := \det A = a_{11}a_{22} - a_{12}a_{21}$. It is well-known that necessary and sufficient conditions for the stability of $z^2 + a_1 z + a_2 = 0$ are (Jury (1974); Goodwin and Sin (1984)): $a_2 < 1$ plus $1 + a_1 + a_2 > 0$ and $1 - a_1 + a_2 > 0$. Since a_2 is a non-linear, non-convex function of the original parameters (the Hessian of a_2 with respect to the four a_{ij}s is not positive-definite), the stability region is non-convex. Moreover, it is unbounded. However, we can state the following necessary *linear* constraint in this case, resulting from the substitution of $a_2 < 1$ in the other two inequalities:

$$-2 < a_{11} + a_{22} < 2 \quad \text{or} \quad |\text{trace}\, A| < 2. \qquad \square$$

Definition 1. Let $C(m, R)$ denote the set of complex numbers within or at a circle with

real-values centre m and radius R, i.e.

$$C(m, R) := \{z \in \mathbb{C} \mid \|z - m\| \le R;\ m \in \mathbb{R};\ R \in \mathbb{R}^+\}.$$

Lemma 1. Necessary linear conditions (not the sharpest which will be discussed below) such that all roots of the polynomial equation $z^l + a_{l-1}z^{l-1} + \cdots + a_1 z + a_0 = 0$ fall in $C(0, R)$ are

$$|a_i| \le R\binom{l}{i}, \quad \text{for } i = 1, 2, \ldots, l.$$

Proof. Let the roots $z_1, \ldots, z_l \in C(0, R)$, i.e. $|z_i| \le R$. For the ith coefficient a_i one can write

$$|a_i| = \left|\sum_{1 \le k_1 < k_2 < \cdots < k_i \le l} z_{k_1} \cdots z_{k_i}\right|$$

$$\le \sum_{1 \le k_1 < k_2 < \cdots < k_i \le l} |z_{k_1}| \cdots |z_{k_i}|$$

$$\le \sum_{1 \le k_1 < \cdots < k_i \le l} R^i$$

$$= [\text{number of ways to select } i \text{ objects out of } l] \times R^i = R^i\binom{l}{i}. \quad \blacksquare$$

Theorem 1. If all nv poles of (7) are in $C(0, R)$ then $|\text{trace}\, A_1| \le Rnv$ and $|\det A_n| \le R^{nv}$.

Proof. Consider the polynomial following from the expansion of $\det \mathcal{A}(z^{-1}) = 0$

$$z^{nv} + a_{nv-1}z^{nv-1} + \cdots + a_1 z + a_0 = 0$$

Since the roots of this polynomial are in $C(0, R)$, it follows from Lemma 1 that $|a_i| \le R^i\binom{nv}{i}$ for $1 \le i \le nv$. On the other hand, the coefficient a_1 of z in the above equation equals trace A_1 as can be easily checked. Thus

$$|\text{trace}\, A_1| = |a_1| \le R\binom{nv}{1} = Rnv$$

Moreover $|\det A_n| = |a_{nv}| \le R^{nv}$. $\blacksquare$

It will be clarified below that the poles of discrete-time models that emerge from *correctly sampled*, continuous-time systems cannot be situated in $\mathbb{C}^-$, the open left half of the complex plane. Let the underlying continuous-time system have n poles of sufficient physical significance (in the sense that they contribute to the dynamical properties in a non-neglectable way) say $\alpha_i \pm j\beta_i$ with $\beta_i \ge 0$. If a discrete-time equivalent model of the same degree is being used, it is well known that its poles are given by $\exp T(\alpha_i + j\beta_i) = \exp(T\alpha_i)[\cos(T\beta_i) + j\sin(T\beta_i)]$ where T is the sampling period. Let T be selected correctly, then the sampling frequency must necessarily satisfy $2\pi/T \ge 2\max_i \beta_i \ge 0$, i.e. minimally twice as fast as the fastest eigenfrequency of practical relevance. This choice is based on the sampling theorem (e.g. Franklin and Powell (1980)). However, for the choice of the sampling frequency the -3 dB frequency of the system is more important. As the -3 dB point usually will be far beyond $\max_i \beta_i$, one should select the sampling frequency considerably faster than the conservative lower bound mentioned above. Thus, it is perfectly reasonable to asume that the sampling frequency at least satisfies $2\pi/T \ge 4\max_i \beta_i \ge 0$. Consequently, the real parts of the discrete-time poles (i.e. $\Re e(\alpha_i \pm j\beta_i) = \exp(Y\alpha_i)\cos(T\beta_i)$) are situated between 0 and $\exp(T\alpha_i)$, i.e.

$$0 = \exp(T\alpha_i)\cos(\pi/2)$$
$$\le \exp(T\alpha_i)\cos(T\beta_i)$$
$$\le \exp(T\alpha_i)$$

We may therefore conclude that poles of *discrete-time* models that emerge from *correctly sampled*, continuous-time systems cannot be situated everywhere in $C\left(0, \max_i \exp(T\alpha_i)\right)$, but in addition, must be situated in $\mathbb{C}\backslash\mathbb{C}^-$. This additional knowledge can be translated into inequality constraints on the parameters too.

Lemma 2. A necessary (but not sufficient, except for degree $l \le 2$) condition for the zeros of $\zeta^l + d_{l-1}\zeta^{l-1} + \cdots + d_1\zeta + d_0$ ($\zeta \in \mathbb{C}$, $d_i \in \mathbb{R}$) to have positive real parts, is:

$$d_i \ge 0 \quad (i = 1, 2, \ldots, l).$$

Proof. See e.g. Casti (1977). $\blacksquare$

Theorem 2. If all poles of the general model (7) have non-negative real parts, the following inequalities hold: trace $A_1 \le 0$ (linear) and $(-1)^{nv}\det A_n \ge 0$. In case (7) is scalar ($v = 1$), its coefficients a_i satisfy $(-1)^i a_i \ge 0$, $1 \le i \le n$.

Proof. Since by assumption the zeros of the characteristic polynomial (8) have non-negative real parts, all zeros of the associated equation $(-z)^{nv} + a_{nv-1}(-z)^{nv-1} + \cdots + a_1(-z) + a_0 = 0$ will have non-positive real parts. Application of Lemma 2 yields that $(-1)^i a_i \ge 0$, for all $i = 1, 2, \ldots, nv$. In particular, $-\text{trace}\, A_1 = -a_1 \ge 0$ and $(-1)^{nv}\det A_n = (-1)^{nv}a_{nv} \ge 0$. $\blacksquare$

From here on, we will consider the derivation of *linear* inequalities on the parameters of (7) for

the more general situation in which it is known that all poles of (7) are in circle $C(m, R)$, $m \in \mathbb{R}$, $R \geq 0$. This formulation has considerable practical relevance. Trivially, if one knows that the process is stable, $m = 0$ and $R = 1$ can be selected. If the magnitude of the largest pole is known to be below $R_{\max}$, one can set $m = 0$ and $R = R_{\max}$. If a lower bound for the damping ratio of a stable oscillatory process is known, the logarithmic spirals that embrace the possible locations of the relevant complex pole pairs can be covered by $C(m, 1 - m)$ for suitable $m > 0$. Furthermore, any combination of a number of circle requirements on the process poles can be accommodated with the intersection of the sets of constraints that each circle requirement will be shown to introduce. Of course the "sound poles" constraints (Theorem 2) can be added too.

It is convenient to study the $C(\frac{1}{2}, \frac{1}{2})$ case first, since the associated parameter constraints are relatively easily derived and allow a simple generalization to the $C(m, R)$ case. Consider the Möbius-transformation $z(s) = 1/(1 - s)$ which maps the imaginary axis in the s-domain onto $\partial C(\frac{1}{2}, \frac{1}{2}) \backslash \{0\}$ since $\|z(s = j\omega) - \frac{1}{2}\| = \frac{1}{2}$ for real ω and (due to continuity), $\mathbb{C}^-$ into this circle. This means that the roots of (7) are at or within $C(\frac{1}{2}, \frac{1}{2})$ iff the roots of the Möbius transformed equation (7) are not in $\mathbb{C}^+$. Substitution of $z^{-1} = 1 - s$ in (7) yields $\mathcal{A}(1 - s) = I + A_1(1 - s) + A_2(1 - s)^2 + \cdots + A_n(1 - s)^n$. Binomial expansion shows that the coefficient matrix D_k corresponding to the power s^k is

$$D_k := (-1)^k \times \left[A_k\binom{k}{k} + A_{k+1}\binom{k+1}{k} + \cdots + A_n\binom{n}{k}\right], \quad (0 \leq k \leq n), \tag{8}$$

where we have set $A_0 := I$. These matrix coefficients constitute a polynomial matrix $\mathcal{D}(s) := D_0 + D_1 s + D_2 s^2 + \cdots + D_n s^n$. Based on the above established relationship, the roots of (7) strictly are in $C(\frac{1}{2}, \frac{1}{2})$ iff all roots of $\det \mathcal{D}(s) = 0$ have negative real part. Necessary and sufficient stability conditions, characterized in terms of the coefficients of $\mathcal{D}$ (e.g. Shieh and Sacheti (1978); Shieh *et al.* (1987)) for the general, multivariable case, invariably lead to non-linear constraints not suitable for our needs. However, in case all A_k in (7) are either upper or lower *triangular* matrices (and thus all D_k in (8) being triangular too) it is possible to describe a smallest, linearly bounded, convex hull of the admissible parameter region. The first observation is that the non-zero off-diagonal terms may have any value. They do not affect the position of the poles at all, as can be seen from $\det \mathcal{A}(z^{-1}) = 0$. Secondly, the analysis simplifies to the question whether the poles of *each* of the v scalar systems of the form (corresponding to each diagonal term of $\mathcal{A}(z^{-1})$):

$$a(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}, \tag{9.1}$$

are strictly in $C(\frac{1}{2}, \frac{1}{2})$ (for simplicity we have dropped the indices of the individual polynomials). Equivalently, we could ask whether all poles of

$$d(s) := d_0 + d_1 s + d_2 s^2 + \cdots + d_n s^n, \qquad d_k := (-1)^k \sum_{l=k}^{n} a_l \binom{l}{k}, \tag{9.2}$$

are in $\mathbb{C}^-$. Using Lemma 2, we may conclude that $d_k \geq 0$ for $1 \leq k \leq n$. Consequently, *necessary* conditions for the zeros of (9.1) to be in $C(\frac{1}{2}, \frac{1}{2})$ are:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 0 & -1 & -2 & -3 & \cdots & -\binom{n-1}{1} & -\binom{n}{1} \\ 0 & 0 & 1 & 3 & \cdots & \binom{n-1}{2} & \binom{n}{2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & (-1)^{n-1} & \\ 0 & 0 & 0 & 0 & \cdots & 0 & (-1)^n \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

or, briefly, $\mathcal{R} \times a \geq 0$ where $a := [1, a_1, \ldots, a_n]^T$ is the coefficient vector and $\mathcal{R}$ is an $n \times n$ upper triangular matrix with $[\mathcal{R}]_{ij} = (-1)^i\binom{j}{i}$ for each non-zero (i, j)th element $(0 \leq I \leq n; 0 \leq j \leq i)$. In other words, the triangular part simply equals a sign-corrected version of Pascal's triangular expansion. The recurrence in (10) allows simple expansion, whereas all lower-dimensional cases remain available. The $n + 1$ equations in (10) actually describe a convex set in the n-dimensional coefficient space, known as *simplex*. This simplex is spanned by all linear combinations of $n + 1$ vertices. These vertices can be read from the columns of $\mathcal{R}$, skipping each first element because of the leading constant in the parameter vector. This can be seen from the fact that the jth contribution $(0 \leq j \leq n)$ corresponds to a parameter vector with ith element $(-1)^i\binom{j}{i}$

for $1 \le i \le j$ and zero for $j+1 \le i \le n$. Consequently, it represents a system $a(z^{-1}) = 1 - \binom{j}{1}z^{-1} + \binom{j}{2}z^{-2} + \cdots + (-1)^j\binom{j}{j}z^{-j} = (1 - z^{-1})^j$, having $(n-j)$ poles at $z = 0$ and j at $z = 1$. Since these poles are at the edge of $C(\frac{1}{2}, \frac{1}{2})$, the associated jth parameter vector must be on the boundary of the admissible parameter region. Since this is true for all j, the set of linear constraints (10) constitutes the *smallest* convex hull of the admissible region.

We will now discuss the simple relation to the general case where the poles of (9.1) are in $C(m, R)$. Note that $C(m, R)$ is the image of $C(\frac{1}{2}, \frac{1}{2})$ under the linear mapping $q(z) := \gamma z + \beta$ where $\gamma := 2R$ and $\beta := m - R$. Therefore, if the poles of $a(q) = 0$ are in $C(m, R)$ this implies that the poles of $a(\gamma z + \beta) = 0$ are in $C(\frac{1}{2}, \frac{1}{2})$. Since we have derived necessary conditions (10) for that property, we expand $q^n a(q)$ in z, introducing coefficients c_i:

$$
\begin{aligned}
&(\gamma z + \beta)^n a(\gamma z + \beta) \\
&\quad = (\gamma z + \beta)^n + a_1(\gamma z + \beta)^{n-1} + \cdots + a_n \\
&\quad = \gamma^n[z^n + c_1 z^{n-1} + \cdots + c_{n-1}z + c_n].
\end{aligned}
$$

Application of (10) yields $\mathscr{R} \times c \geqq 0$ where $c := [1, c_1, \ldots, c_n]'$. On the other hand it is easily seen that the kth coefficient $(0 \le k \le n)$ in c equals:

$$
c_k = \left[\beta^k\binom{n}{k} + \beta^{k-1}\binom{n-1}{k-1}a_1 + \cdots + \binom{n-k}{0}a_k\right] \Big/ \gamma^k,
$$

stating that $c = \mathscr{L} \times a$ where $\mathscr{L}$ is a lower triangular matrix with $[\mathscr{L}]_{ij} = \beta^{i-j}\binom{n-j+1}{i-j}/\gamma^{i-1}$ for non-zero elements. In conclusion, we have shown the following.

Theorem 3. If the poles of $1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n}$ are in $C(m, R)$ then the following *linear* constraints on the coefficients are satisfied:

$$
\mathscr{R} \times \mathscr{L} \times [1, a_1, a_2, \ldots, a_n]' \geqq 0, \qquad (11)
$$

where the non-zero elements of the right and left triangular matrices $\mathscr{R}$ and $\mathscr{L}$ are defined as:

$$
[\mathscr{R}]_{ij} = (-1)^i\binom{j}{i},
$$

$$
[\mathscr{L}]_{ij} = (m - R)^{i-j}\binom{n-j+1}{i-j}(2R)^{n-i+1}. \quad \blacksquare
$$

Note that since $\mathscr{L}$ is linear, the admissible region given by (11) remains a simplex. Likewise, its $n+1$ spanning vertices can be read from the columns of $(\mathscr{R} \times \mathscr{L})$ skipping each first element. Consequently, since the systems corresponding to the spanning vertices have all their poles at $m \pm R$, i.e. at the edge of $C(m, R)$, (11) constitutes the *smallest* convex hull of the admissible parameter region corresponding to scalar systems having all poles in $C(m, R)$.

3.2.3. *Example* 3: *Stability constraints for scalar AR system*. It is straightforward to derive conditions for (asymptotic) stability of (9.1) from the (strict) inequalities (11). Since the unit circle is the discrete-time stability boundary, we set $m = 0$ and $R = 1$, yielding

$$
[\mathscr{L}]_{ij} = (-1)^{i-j}\binom{n-j+1}{i-j}2^{n-i+1}.
$$

For example, $\mathscr{R} \times \mathscr{L}$ becomes, for $n = 3$:

$$
\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & -1 \end{bmatrix}
\begin{bmatrix} 8 & 0 & 0 & 0 \\ -12 & 4 & 0 & 0 \\ 6 & -4 & 2 & 0 \\ -1 & 1 & -1 & 1 \end{bmatrix}
= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 3 & 1 & -1 & -3 \\ 3 & -1 & -1 & 3 \\ 1 & -1 & 1 & -1 \end{bmatrix}.
$$

So, it is seen from the coefficients in the first and the last matrix that linear $C(\frac{1}{2}, \frac{1}{2})$ and stability constraints for the third-order system $y_k + a_1 y_{k-1} + a_2 y_{k-2} + a_3 y_{k-3} = 0$ are:

$$
\begin{aligned}
1 + a_1 + a_2 + a_3 &\geqq 0; & 1 + a_1 + a_2 + a_3 &> 0, \\
-a_1 - 2a_2 - 3a_3 &\geqq 0; & 3 + a_1 - a_2 - 3a_3 &> 0, \\
a_2 + 3a_3 &\geqq 0; & 3 - a_1 - a_2 + 3a_3 &> 0, \\
-a_3 &\geqq 0; & 1 - a_1 + a_2 - a_3 &> 0.
\end{aligned}
$$

The induced simplices are shown in Fig. 4 together with their spanning vertices. For example, (0, 0, 0), (−1, 0, 0), (−2, 1, 0) and (−3, 3, −1) span the $C(\frac{1}{2}, \frac{1}{2})$ simplex. Jury (1974) and Peeters (1987) have treated the stability case directly. However, Theorem 3 is more transparent due to the $\mathscr{R} \times \mathscr{L}$ decomposition and it is not restricted to stability. $\square$

3.2.4. *Constraints on stationary gains*. Often not only (asymptotic) stability of the plant is known beforehand, but also admissible intervals, or at least signs, for the several static gains in such a stable multivariable process. As argued in Section 2, "best" identified models often reproduce low-frequency characteristics poorly, if at all. The significance of *a priori* knowledge must, therefore, not be underestimated. It will be shown in the sequel that information about the (sign of the) static gain will introduce a dramatic increase of overall model quality. This can be crucial to satisfactory controller design. For instance, Nemir and Kashyap (1986) show

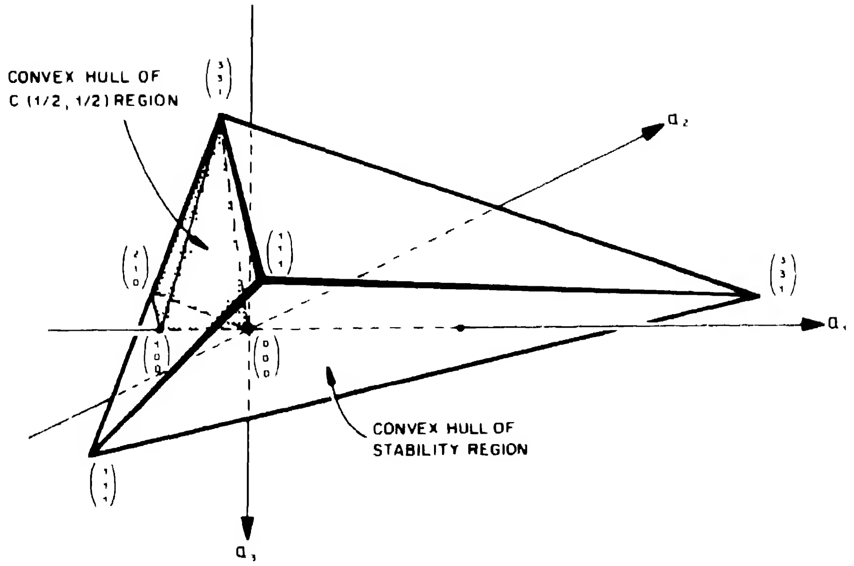


FIG 4 Convex hulls of the stability and $C(1/2, 1/2)$ regions

that in the case of a minimax one-step-ahead variance criterion, *open-loop* operation of a stable process is optimal when the model suffers from uncertain low-frequency characteristics.

When the deterministic part of the process has an asymptotically stable transfer function (6), the stationary gain matrix $\mathscr{G}$ equals $H(1) = [\mathscr{A}(1)]^{-1}\mathscr{B}(1)$. Now, suppose $\mathscr{G}$ is known to be between a lower bound $\mathscr{G}^-$ and an upper bound $\mathscr{G}^+$, i.e. $\mathscr{G}^- \leq \mathscr{G} \leq \mathscr{G}^+$ in an *element-wise* fashion, allowing $\pm\infty$. If $\mathscr{A}(1) \geq 0$, one may conclude that $\mathscr{A}(1)\mathscr{G}^- \leq \mathscr{B}(1) \leq \mathscr{A}(1)\mathscr{G}^+$, or equivalently:

$$\begin{aligned}(I + A_1 + \cdots + A_n) \times \mathscr{G}^- \\ \leq (B_1 + \cdots + B_m) \\ \leq (I + A_1 + \cdots + A_n) \times \mathscr{G}^+. \quad (12)\end{aligned}$$

Note that $\mathscr{A}(1) \geq 0$ (element-wise) is not a consequence of asymptotic stability of $\mathscr{A}$ in the general multivariable case, even if $\mathscr{A}$ is triangular as we assumed before. However, it is true if the asymptotically stable $\mathscr{A}$ is of *diagonal shape*. This is a necessary consequence of the fact that in that case each ith scalar system $a_i(z^{-1})$, read from the ith diagonal element $\mathscr{A}_{ii}$ of $\mathscr{A}$, has to be asymptotically stable. Then the first equation in (10) implies (cf. Example 3)

$\mathscr{A}_{ii}(1) = a_i(1) = \sum_{i \leq k \leq n} a_k^i > 0$ for all i, where a_k^i is the kth coefficient of a_i. For the popular class of FIR (Finite Impulse Response) modelling, where the autoregressive model part is missing ($n = 0$), additional opportunities arise because of the straightforward, linear relation with the successive discrete-time step response matrices $\mathscr{S}(kT)$, $k \geq 1$. In practice, often some idea about

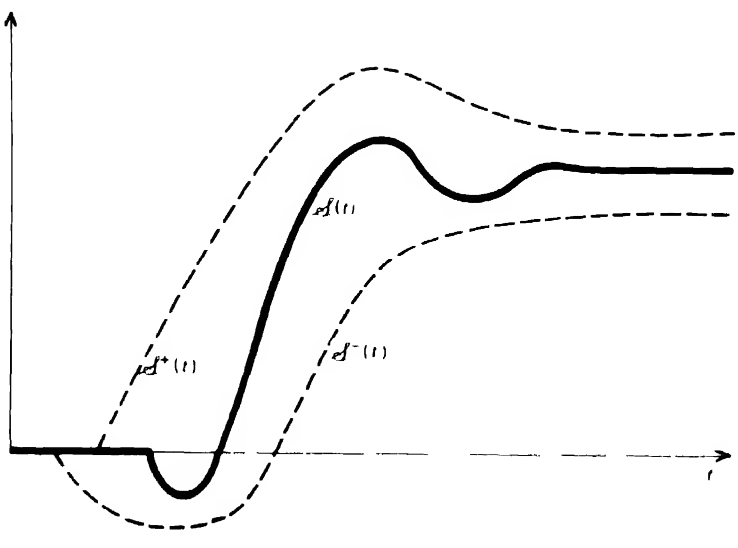


FIG 5. For FIR models a 'tube' of acceptable step responses can be easily translated into linear Grey-Box model constraints

a "tube of *possible* responses", bounded by marrices $\mathcal{S}^+(kT)$ and $\mathcal{S}^-(kT)$ at discrete-time index k, is established over time. This incorporates aspects as dead-times, (non)minimum phase properties, rise/setting time and overshoot (cf. Fig. 5) yielding for $k \geq 1$:

$$\mathcal{S}^-(kT) \leq \sum_{i=1}^{\min(k,m)} B_i \leq \mathcal{S}^+(kT). \quad (13)$$

4 BAYESIAN INFERENCE OF GB PARAMETER DISTRIBUTION USING PHYSICAL KNOWLEDGE

4.1. *Introduction*

In the previous section we have defined a GB model class (4) with GB parameter set (3), containing physically consistent candidate model parameters θ. It was shown how important physical information could be translated in the inequality form (5), see Theorem 2 and equations (11)–(13). In this Section we will adopt a Bayesian approach and regard the unknown parameter estimator of θ to be a random variable (r.v.), denoted $\hat{\Theta}$. Using Bayesian inference, we will construct the following conditional *posterior* p.d.f. for a realization $\hat{\Theta} = \theta$:

$$f_{\hat{\Theta}|Z,\Omega}(\theta \mid Z = z, \hat{\Theta} \in \Omega), \quad (14)$$

which is conditioned not only upon the data stack Z ($Z = z$ denoting observations) but also on the probabilistic Grey-Box statement $\hat{\Theta} \in \Omega$. Hereby we express that *any* realization θ of $\hat{\Theta}$ must be consistent with the *a priori* given knowledge, i.e. $\theta \in \Omega$. Our aim will be to construct (14) from the likelihood Z and a prior p.d.f. which is associated with the GB model class.

4.2. *Bayesian inference in Grey-Box estimation*

The rule of Bayes plays a central role in Bayesian statistics, e.g. Box and Tiao (1973) or Press (1989). Let $\Pr(A \mid B) := \Pr(AB)/\Pr(B)$ denote the conditional probability of the event A, given the occurrence of event B. Since $\Pr(AB) := \Pr(A$ and $B)$ of course equals $\Pr(BA) = \Pr(B$ and $A)$ it follows directly that for $\Pr(B) > 0$

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \times \Pr(A)}{\Pr(B)},$$

which is known as *Bayes' rule*. Setting $A = \hat{\Theta} \mid \Omega$ and $B = Z$, the rule can be used to relate the *posterior* p.d.f. (14) of the s.v. $(\hat{\Theta} \mid Z, \Omega) := (\hat{\Theta} \mid \Omega) \mid Z$, namely

$$f_{\hat{\Theta}|Z,\Omega}(\theta \mid Z = z, \hat{\Theta} \in \Omega) = \frac{f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta) \times f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega)}{\int_{\Omega} f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta) \times f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega)\, d\theta}, \quad (15)$$

to the *prior* p.d.f. $f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega)$ (which equals zero for $\theta \notin \Omega$) and to the *likelihood* $f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta)$, with data z, given prior parameter estimate $(\hat{\Theta} \mid \Omega) = \theta$. The integral in the denominator of (15) is providing consistency with $\Pr((\hat{\Theta} \mid Z, \Omega) \in \Omega) = 1$. This is crucial for the derivation of one of the constrained estimators in the next section. However, for our needs it is not necessary that the prior or the likelihood be normalized. Figure 6 illustrates how the combination of the above introduced elements is organized.

An appealing feature is that the likelihood occurs explicitly in the calculation of (15), allowing for calculation of conventional results as well. First we will consider the likelihood. Given are N output vectors y_k measured at successive time instants and stacked in a measured data matrix Y^N. In addition, a number of known input vectors u_k and noise realizations ξ_k (assumed to be correctly reconstructed or measured) up to time index N are in matrix stacks U^N and Ξ^N. We assume that the output is generated by a 'process' (with true parameter θ^*) from a general class of linearly parametrized regression structures

$$y_k = \mathcal{F}_k(Z^{k-1}) \times \theta + \xi_k,$$
$$Z^{k-1} := [Y^{k-1} \mid U^{k-1} \mid \Xi^{k-1}].$$

For convenience, θ is considered to be a $\eta := vp$ dimensional vector, consisting of the columns of

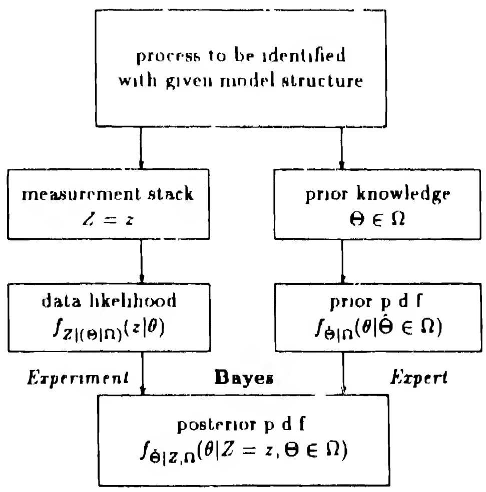


FIG. 6 The Bayesian creation of the posterior from prior and likelihood.

the earlier introduced $v \times \rho$ parameter matrix θ. It is an alternative representation for (2), which includes the ARMAX "process" (1). The successive residues ξ are assumed to be *independent*, but not necessarily identically distributed v-dimensional s.v.s. Then, given this independence, N repetitions of Bayes' rule applied to the likelihood of Y^N given the regressor Z^{N-1} and parameter θ will result in (Goodwin and Payne, 1977):

$$f_{y^N}(Y^N \mid Z^{N-1}, \theta) = \prod_{k=1}^{N} f_{\xi_k}(\xi_k(\theta) \mid Y^{k-1}, Z^{k-1}, \theta),$$

where $\xi_k(\theta) := y_k - \mathcal{F}_k(Z^{k-1}) \times \theta$ is the *prediction error*. We will restrict ourselves here to the case where the white noise sequence $\{\xi_k\}$ is *normally distributed* with common covariance matrix Σ and mean vector $\mu = 0$, i.e. $\xi_k \sim \mathcal{N}(0, \Sigma)$. Then the above likelihood is Gaussian.

$$f_{y^N}(Y^N \mid Z^{N-1}, \theta) = \frac{\exp\left\{-\frac{1}{2}\sum_{k=1}^{N} \xi_k^T \Sigma^{-1} \xi_k\right\}}{[(2\pi)^v \det \Sigma]^{N/2}} \tag{16}$$

However, other choices for the likelihood can be made when applicable, e.g. from the assumption of a uniformly distributed white noise sequence, which case has considerable practical relevance. The other branch in Fig. 6 defines the prior p.d.f. Obviously, this is chosen to be zero outside the GB set Ω, see (3). The choice of the shape within Ω often is a subjective one. We will focus primarily on a *prior* that is uniformly distributed on Ω:

$$f^{\text{uni}}_{\hat\Theta|\Omega}(\theta \mid \hat\Theta \in \Omega) = \begin{cases} 1\Big/\int_\Omega \mathrm{d}\theta & \text{for } \theta \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Using this p.d.f. definition it is necessary that the integral exists, i.e. Ω must be bounded. However, with this uniform prior, the *posterior* p.d.f. becomes *a* to Ω "restricted" likelihood:

$$f^{\text{uni}}_{\Theta|Z,\Omega}(\theta \mid Z = z, \hat\Theta \in \Omega) = \begin{cases} f_{Z|(\hat\Theta|\Omega)}(z \mid \theta)\Big/\int_\Omega f_{Z|(\hat\Theta|\Omega)}(z \mid \theta)\,\mathrm{d}\theta & \text{for } \theta \in \Omega \\ 0 & \text{elsewhere,} \end{cases} \tag{17}$$

see (15). Limiting arguments show that (17) is also valid when Ω is unbounded if the integral in the denominator exists. Contrary to common belief, the uniform prior is usually *informative* on Ω (it thus adds knowledge, although it is constant on Ω) but some aspects of it are not fully satisfactory (Peterka (1981); Bard (1974); Box and Tiao (1973)). In our philosophy, the prior distribution must be made as informative as can be, which is usually not the case with a uniform prior. However, a uniform prior is attractive since precise ideas about the likeliness of each model parameter in Ω are often lacking. Moreover, it allows of analytical derivation of the constrained estimators.

An alternative, and useful novel prior, allowing of prior model discrimination within Ω, is a (continuous) *prior* f^{pl}, *Piece-wise Linearly* distributed on Ω. Here Ω has the form (5) and is assumed *bounded*. This prior is chosen to be zero outside Ω and also at its edge $\partial\Omega$, and is given some global, positive maximum M (normalizing constant) at some fixed interior parameter θ^0, which is considered *a priori* to be most likely. Furthermore, at the line segment from any parameter θ^e on $\partial\Omega$ to $\theta^0 \notin \partial\Omega$, the prior is increasing linearly from 0 to M

$$f^{\text{pl}}_{\hat\Theta|\Omega}((1-\lambda) \times \theta^e + \lambda \times \theta^0 \mid \hat\Theta \in \Omega)) = \lambda \times M, \quad 0 \le \lambda \le 1,\ \theta^e \in \partial\Omega. \tag{18}$$

This definition is consistent since the interiors of all such line segments are disjoint and in Ω because of the convexity of Ω. If Ω is unbounded (this is the typical case in applications) we could add bounding constraints to (3), which can always be chosen such that the integral of the likelihood over the ignored part of Ω is neglectable. Much more attractive (since it is easier and more flexible) is an extension of the *pl* prior with a *uniform* component in the unbounded case. To this end, consider the ith *facet* $\partial\Omega_i$ (with $1 \le i \le \zeta$, where ζ is he number of constraints) on the edge $\partial\Omega$ of Ω, which is defined as the admissible subset with active ith constraint: $\partial\Omega_i := \{\theta \in \Omega \mid a_i^T\theta = b_i\}$. Next, a polyhedron $\Omega_i \subset \Omega$ can be defined by forming all possible linear combinations of $\theta^e \in \partial\Omega_i$ with θ^0, i.e.

$$\Omega_i = \{\theta \in \Omega \mid \theta = (1-\lambda)\theta^e + \lambda\theta^0; 0 \le \lambda \le 1; \theta^e \in \partial\Omega_i; \theta^0 \in \Omega\backslash\partial\Omega\}.$$

Note that Ω_i is a *cone* with top θ^0 and base hyperplane $a_i^T\theta = b_i$, containing the facet $\partial\Omega_i$. The *plu*riform (Piece-wise Linear with possibly Uniform component) prior is defined as (cf. Fig. 7)

$$f^{\text{plu}}_{\hat\Theta|\Omega}(\theta \mid \hat\Theta \in \Omega) = \begin{cases} 0 & \text{if } \theta \notin \Omega \\ (a_i^T\theta - b_i)/(a_i^T\theta^0 - b_i) & \text{if } \theta \in \Omega_i \quad \text{for some } i\ (1 \le i \le \zeta) \\ 1 & \text{otherwise} \quad \text{(i.e. a uniform component).} \end{cases}$$

Another choice, well-known from Bayesian

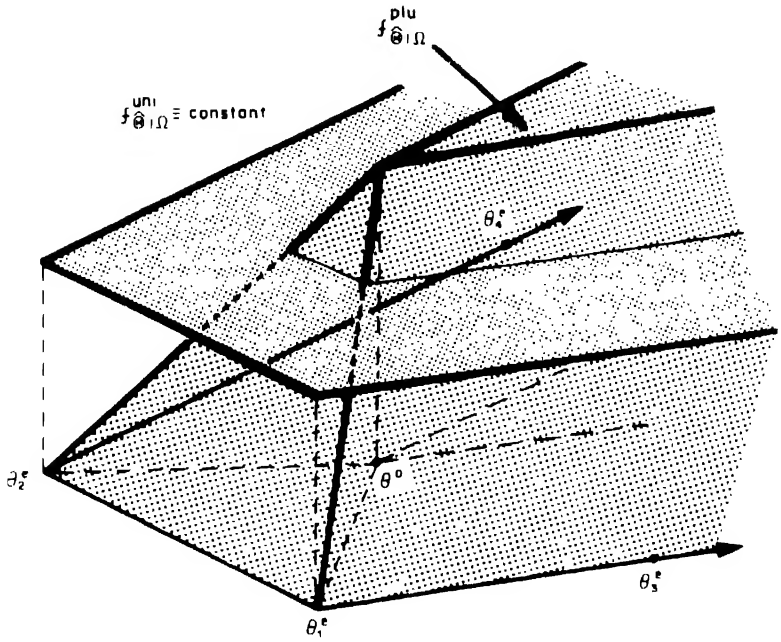


FIG. 7. UNIform and PLUriform prior example for unbounded parameter set Ω.

inference, is a Gaussian prior with mean vector θ^0 and covariance matrix Π. In the GB context one can consider a *Restricted Gaussian* prior:

$$f^{rg}_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega) = \begin{cases} \dfrac{C \exp\{-\frac{1}{2}(\theta-\theta^0)'\Pi^{-1}(\theta-\theta^0)\}}{\sqrt{(2\pi)^{\nu_p}\det\Pi}} & \text{for } \theta \in \Omega \\ 0 & \text{elsewhere,} \end{cases} \tag{19}$$

where $C \geq 1$ is the p.d.f. normalization constant. Note that Ω may be unbounded. However, a practical drawback of this prior is the non-trivial specification of the prior parameter covariance matrix Π, with the possible exception of FIR modelling.

5. BAYESIAN MAP AND AAP ESTIMATORS

5.1. *Introduction*

In the previous section it was explained how the posterior (15) can be calculated. In addition, some specific choices of the prior and the likelihood have been presented. Once a posterior has been made available, it presents information about the "likeliness" of the very many potential models, based on data as well as prior knowledge. However, this information does not imply directly which model should be preferred. Thus, since for practical reasons we need one particular model, a criterion should be chosen in order to define what is "best". In this section we will consider a Maximum *A Posteriori* (MAP) and an Average *A Posteriori* (AAP) estimator. Both are Bayesian analogies of standard (non-Bayesian) estimation concepts.

5.2. *Definition of the MAP and CML estimator*

The conventional Maximum Likelihood (ML) estimator selects the parameter that maximizes the data likelihood (Eykhoff, 1974; Goodwin and Payne, 1977):

$$\hat{\Theta}_{ML} := \arg\max_{\theta \in \Psi} f_{Z|\hat{\Theta}}(z \mid \theta),$$

over an admissible parameter set Ψ which is usually unconstrained, $\Psi = \mathbb{R}^{\nu_p}$. We define a Maximum *A Posteriori* (MAP) estimator using the same type of criterion (Box and Tiao, 1973; Ljung and Söderström, 1983; Goodwin and Sin, 1984) where the posterior replaces the likelihood:

$$\hat{\Theta}_{MAP} := \arg\max_{\theta \in \Omega} f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta) \times f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega). \tag{20}$$

If the prior in (20) is *uniformly* distributed on Ω (cf. (17)), this "Bayesian" estimator is equivalent to an ML estimator that is constrained with respect to $\Psi = \Omega$. We will therefore refer to it as MAP/*un* or *Constrained Maximum Likelihood* (CML) estimator. Note that

$$\hat{\Theta}_{CML} := \begin{cases} \hat{\Theta}_{ML} & \text{if } \hat{\Theta}_{ML} \in \Omega \\ Proj_{\partial\Omega}(\hat{\Theta}_{ML}) & \text{otherwise,} \end{cases} \tag{21}$$

where the latter expression is the most likely, in general non-orthogonal projection on $\partial\Omega$ of the *unconstrained ML* estimator, cf. Goodwin and

Sin (1984). However, when the prior is nonuniformly distributed, this relation is more complicated. We briefly treat the interesting case where the prior is *piecewise linearly* distributed (cf. (18)) on a *bounded*, polyhedral set Ω of the form (4), with *Gaussian* likelihood (16) and linearly parametrized regression model. The associated estimator seems to be more attractive than the CML estimator, since it is composed of a (in general) more informative prior. Moreover, it will never be on $\partial\Omega$. This is attractive as it is unlikely or even impossible that model parameters are on $\partial\Omega$ when Ω follows from reliable knowledge. With the piece-wise linear prior assumption, it is attractive to decompose Ω in a finite number of almost disjoint cones Ω_i which have the vertex $\theta^0 \in \Omega\backslash\partial\Omega$ as common top, cf. (19). Moreover, the prior behaves strictly linear on each cone. Each linear prior component follows from the fact that it is positive (the value is not crucial and can, e.g. be chosen unity for all indices) at the inner vertex θ^0 and zero at the vertices on $\partial\Omega$.

Theorem 4. If the maximizing argument of (20) is an *interior* point of some cone $\mathscr{P} \subset \Omega$ having base facet $\Omega \cap \mathscr{P} = \{\theta \mid \alpha^T\theta + \beta = 0\}$ and top θ^0, we obtain for $\lambda := \alpha^T\hat{\Theta}^{pl}_{MAP} + \beta$:

$$\hat{\Theta}^{pl}_{MAP} = \left[\sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}\mathscr{F}_k\right]^{-1} \times \left[\alpha/\lambda + \sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}y_k\right]. \quad (22)$$

Proof. If the maximizing argument of (20) is in the interior of one of the cones that constitute Ω, say P, it follows from:

$$\hat{\Theta}^{pl}_{MAP} = \arg\max_{\theta\in\mathscr{P}} \left[(\alpha^T\theta + \beta) \times \exp\left\{-\tfrac{1}{2}\sum_{k=1}^{N}(y_k - \mathscr{F}_k\theta)^T\Sigma^{-1} \times (y_k - \mathscr{F}_k\theta)\right\}\right],$$

where $\alpha^T\theta + \beta$ is the non-negative linear prior behaviour on P (zero on ∂P). Equating the derivative of the posterior to zero yields the necessary condition for the unconstrained case:

$$\exp\left\{-\tfrac{1}{2}\sum_{k=1}^{N}(y_k - \mathscr{F}_k\theta)^T\Sigma^{-1}(y_k - \mathscr{F}_k\theta)\right\} \times \left\{\alpha - (\alpha^T\theta + \beta) \times \sum_{k=1}^{N}[\mathscr{F}_k^T\Sigma^{-1}\mathscr{F}_k\theta - \mathscr{F}_k^T\Sigma^{-1}y_k]\right\} = 0,$$

from which it follows, introducing

$$\varphi := \sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}y_k \quad \text{and} \quad \mathscr{F} := \sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}\mathscr{F}_k:$$

$$\theta = \mathscr{F}^{-1}[\alpha/(\alpha^T\theta + \beta) + \varphi] = \mathscr{F}^{-1}[\alpha/\lambda + \varphi],$$

stating the desired result. The real-valued quantity λ is the non-negative solution (since $\hat{\Theta}^{pl}_{MAP}$ must be admissible) to the quadratic problem

$$\lambda^2 - (\alpha^T\mathscr{F}^{-1}\varphi + \beta) \times \lambda - \alpha^T\mathscr{F}^{-1}\alpha = 0. \quad \blacksquare$$

Note that the solution (22) for $\alpha = 0$ reduces to the standard ML case, here resulting in a Weighted Least Squares (WLS) estimator

$$\hat{\Theta}_{WLS} := \left[\sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}\mathscr{F}_k\right]^{-1}\left[\sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}y_k\right], \quad (23.1)$$

since Σ is assumed known (Goodwin and Payne, 1977). We can thus interpret (22) as a shifted (corrected) WLS estimator. When it is not an interior point of one of these cones, it must be on some *interior* intersection of at least two cones, since the posterior equals zero on $\partial\Omega$. An algorithm can be given which systematically locates the maximum. Apart from some remarks, we will discuss it elsewhere. It has some relation with the projection algorithm needed in the CML case (uniform prior) for Gaussian likelihood and polyhedral Ω, which is explained in detail by Liew (1976) and Peeters (1987). As in that case the parameter conditioned with respect to data is a Gaussian distributed s.v. $\hat{\Theta} \mid (Z = z) \sim \mathcal{N}(\mu, \Pi)$ where (see (23.1))

$$\mu := \mathscr{E}\hat{\Theta}_{WLS};$$

$$\Pi := \operatorname{var}\hat{\Theta}_{WLS} = \left[\sum_{k=1}^{N} \mathscr{F}_k^T\Sigma^{-1}\mathscr{F}_k\right]^{-1}, \quad (23.2)$$

since the prediction errors are assumed to be linearly dependent on θ. In the other cases, we consider the normal assumption a reasonable one. In order to solve the CML and MAP (for piecewise linear prior) problems it is very convenient to introduce the linear transformation $\bar{\theta} := \mathscr{R}^{-T} \times (\theta - \mu)$ as it (stochastically) orthogonalizes the transformed parameter. Here $\mathscr{R}$ is a Choleski factor of the above covariance matrix Π, i.e. $\Pi = \mathscr{R}^T\mathscr{R}$. The above notions are for the *transformed*, two-dimensional situation illustrated in Fig. 8. It shows a situation with transformed mean $\bar{\mu} \in \bar{\Omega}$, respectively $\bar{\mu} \notin \bar{\Omega}$ (the circles represent equidensity curves of the transformed Gaussian s.v.). In addition, the CML projection onto $\partial\bar{\Omega}$ is shown for the non-admissible mean. Figure 8 also offers some insight into the MAP with piecewise linear (pl)

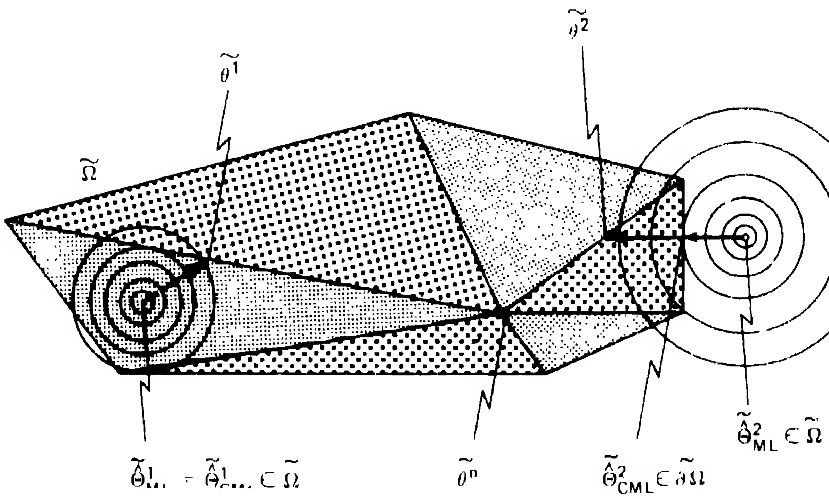


FIG. 8. ML, CML and MAP/*plu* estimators in transformed two-dimensional space (using orthogonalized Gaussian Likelihood).

prior (with centre θ^0) estimator. In both situations, MAP/*pl* is in or on the edge of the cone (in this case a two-dimensional simplex, i.e. triangle) in which the unconstrained ML or its CML projection is located. When it is within this simplex, the transformed result is on the line segment that starts off with the CML estimator and which is perpendicular to and directed off the edge, cf. Theorem 4. A detailed discussion of the merits of the novel MAP/*plu* estimator and its properties together with an efficient algorithm will be published elsewhere.

5.3. *Definition of AAP and CMV estimator*

The conventional Minimum Variance (MV) estimator is defined as the following conditional expectation

$$\hat{\Theta}_{MV} := \int_{\mathbb{R}^{\nu p}} \theta f_{Z|\Theta}(z \mid \theta)\,d\theta = \mathscr{E}[\hat{\Theta} \mid Z = z].$$

This coincides with the ML estimator for a symmetrical likelihood and minimizes the mean parameter error variance (Ljung and Söderström, 1983). We define an Average *A Posteriori* (AAP) estimator using the same type of criterion, replacing the likelihood by the posterior (15):

$$\hat{\Theta}_{AAP} := \mathscr{E}[\hat{\Theta} \mid Z = z, \hat{\Theta} \in \Omega]$$

$$= \frac{\int_\Omega \theta \times f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta) \times f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega)\,d\theta}{\int_\Omega f_{Z|(\Theta|\Omega)}(z \mid \theta) \times f_{\hat{\Theta}|\Omega}(\theta \mid \hat{\Theta} \in \Omega)\,d\theta}.$$

Next we will discuss, in some detail, the analytical calculation of the AAP estimator when the prior is *uniformly distributed* on a polyhedral set $\Omega := \{\theta \in \mathbb{R}^{\nu p} \mid A\theta \le b\}$ Similar to the situation where MAP for *uniform* prior simplified to restricted ML (CML), we can interpret AAP/*un* in the uniform case as (with respect to Ω) constrained MV (CMV) estimation, which equals, cf. (17)

$$\hat{\Theta}_{CMV} = \int_\Omega \theta \times f_{Z|(\hat{\Theta}|\Omega)}(z \mid \theta)\,d\theta \Big/ \int_\Omega f_{Z|(\Theta|\Omega)}(z \mid \theta)\,d\theta. \quad (24)$$

The denominator is needed for p.d.f. normalization. For notational convenience and to stress the generality, we will introduce the abbreviation $X := \hat{\Theta} \mid (Z = z)$. Thus we are interested in

$$\hat{\Theta}_{CMV} = \mathscr{E}[X \mid X \in \Omega] = \mathscr{E}[X \mid AX \le b],$$

where A is a $\zeta \times \eta$ matrix ($\eta := \nu p$) and b is a ζ-vector, following from (5). Observe from (24) that

$$\mathscr{E}[X \mid AX \le b] = \mu + \int_{Ax \le b} (x - \mu) \times f_X(x)\,dx / \Pr(AX \le b), \quad (25)$$

writing $\mu := \mathscr{E}[X]$ for the unconditioned mean. From here, we assume X to be *Gaussian distributed* with mean vector μ and covariance matrix Π, i.e. $X \sim \mathcal{N}(\mu, \Pi)$ as we did in the discussion of the MAP problem. We will first focus on the calculation of $\Pr(AX \le b)$ for the case $\zeta \le \eta$ where A is assumed to have full rank ζ. To this end, define the s.v. $Z := AX$. It is well-known that since Z depends linearly on X, it is also Gaussian with $Z \sim \mathcal{N}(A\mu, A\Pi A^T)$. Let $\mathscr{W}$ be a $\zeta \times \zeta$ Choleski factor of cov(Z) (i.e. $\mathscr{W}^T\mathscr{W} = A\Pi A^T$) then, since $\mathscr{W}^{-T}(Z - A\mu) \sim$

$\mathcal{N}(0, I)$:

$$Pr(AX \le b) = \Pr(Z \le b) = \Pr(\mathcal{W}^{-T}(Z - A\mu) \le \mathcal{W}^{-T}(b - A\mu)) = \prod_{i=1}^{\zeta} \Phi \, [i\text{th element of } \mathcal{W}^{-T}(b - A\mu)], \quad (26)$$

where $\Phi[t] := \Pr(\chi \le t) = \int_{-\infty}^{t} e^{-x^2/2}\,dx/\sqrt{2\pi}$ is the probability function of $\chi \sim \mathcal{N}(0, 1)$. Next we concentrate on the nominator in (25). Let $\mathcal{R}$ be a $\eta \times \eta$ Choleski factor of Π, that is, $\mathcal{R}^T\mathcal{R} = \Pi$. Then, using the linear transformation $y := \mathcal{R}^{-T}(x - \mu)$ with Jacobian det $\mathcal{R}$:

$$\int_{Ax \le b} (x - \mu) f_X(x)\,dx = \frac{\mathcal{R}}{(2\pi)^{\eta/2}} \int_{\Psi} y \times \exp\left(-\frac{y^T y}{2}\right) dy = \mathcal{R}^T \mathscr{E}[Y \mid Y \in \Psi] \times \Pr(Y \in \Psi), \quad (27)$$

where the η-dimensional s.v. $Y := \mathcal{R}^{-T}(X - \mu) \sim \mathcal{N}(0, I)$, and the set Ψ is the transformation of Ω:

$$\Psi := \mathcal{R}^{-T}(\Omega - \mu) = \{y \in \mathbb{R}^\eta \mid A\mathcal{R}^T y \le b - A\mu\}.$$

Although the general case is derived along similar lines (it is however much more involved and will not be treated here), we restrict ourselves here to the case where Ω is defined with a *single constraint* $a^T x \le b$ ($b \in \mathbb{R}$) yielding

$$\Psi = \{y \in \mathbb{R}^\eta \mid c^T y \le d\};$$
$$c := \mathcal{R}a; \quad d := b - a^T\mu.$$

In that case we can write for the ith component Y_i of the vector Y in (27), assuming that ith component c_i of the vector c is not zero:

$$\mathscr{E}[Y_i \mid c^T Y \le d] = \frac{\text{sign}(c_i)}{(2\pi)^{\eta/2}\Pr(c^T Y \le d)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{(d - c_{-i}^T y_{-i})/c_i} y_i \exp\left(-\frac{y_i^2}{2}\right) dy_i \times \exp\left(-\frac{y_{-i}^T y_{-i}}{2}\right) dy_{-i},$$

where we introduced vectors y_{-i} and c_{-i} of reduced dimension $\eta - 1$ by skipping the ith component from y and c, e.g. $c_{-i}^T := [c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_\eta]$. However, a tedious derivation (using the Matrix Inversion Lemma) shows that the right hand side of the above equals

$$\frac{\text{sign}(c_i)}{(2\pi)^{\eta/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}[(y_{-i} - \mu_i)^T \Pi_i^{-1}(y_{-i} - \mu_i) - \sigma_i]\right) dy_{-i},$$

introducing

$$\sigma_i := d^2 \times \left[\frac{\|c_{-i}\|^2}{\|c\|^4} - \frac{1}{c_i^2}\right];$$
$$\Pi_i := [I + c_{-i}c_{-i}^T/c_i^2]^{-1};$$
$$\mu_i := d \times \Pi_i c_{-i}/c_i^2.$$

The above integral expression equals sign $(c_i) \times \exp(\sigma_i/2) \times \sqrt{\det \Pi/2\pi}$, since the integrand is proportional to a Gaussian p.d.f. Moreover, $\det \Pi_i = c_i^2/\|c\|^2$ and we thus conclude to

$$\mathscr{E}[Y_i \mid c^T Y \le d] = \frac{-c_i \times \exp(\sigma_i/2)}{\|c\|\sqrt{2\pi} \times \Pr(c^T Y \le d)},$$

if $c_i \ne 0$, and 0 otherwise as can be directly checked. Finally, using $c = \mathcal{R}a$, $d = b - a^T\mu$ and (26), we substitute the above in (27) and next in (25) to find:

$$\mathscr{E}[X \mid a^T X \le b] = \mu - \frac{\mathcal{R}^T v}{\sqrt{2\pi} \times \|\mathcal{R}a\| \times \Phi[(b - a^T\mu)/\|\mathcal{R}a\|]},$$

where $X \sim \mathcal{N}(\mu, \mathcal{R}^T\mathcal{R})$ and the ith component of the η-dimensional vector v equals 0 if $c_i = 0$ and otherwise $v_i = c_i \times \exp(\frac{1}{2}(b - a^T\mu)^2 (\|c_{-i}\|^2/\|c\|^4 - 1/c_i^2))$. Application to $X = \hat{\Theta} \mid (Z = z)$ with μ and $\Pi = \mathcal{R}^T\mathcal{R}$ from (23.2) results in the CMV estimator for a single constraint (e.g. sign of steady gain)

$$\hat{\Theta}_{CMV} = \mathscr{E}[\Theta \mid Z = z, a^T\hat{\Theta} \le b].$$

Below a simple example is given which illustrates several characteristics shared with the general MAP/un (CML) and AAP/un (CMV) estimators. Next, some properties are highlighted.

5.4. *Example: a CML and CMV estimator in the scalar case*

Consider a *white* stochastic signal $\{X_n \mid n = 1, 2, 3 \cdots\}$ that is uniformly distributed on $[-\mu, \mu]$ where $\mu \in \mathbb{R}^+$. In other words, the X_n are independently, identically distributed uniform r.v.s on $[-\mu, \mu]$. However, the precise value of μ is unknown and has to be estimated. For example, consider a measuring device (having error statistics that are about uniform) of which, e.g. the standard deviation $\mu/\sqrt{2}$ is required. In order to estimate μ, we can consider the realization (outcome, sample) x_n that X_n takes at each time n (i.e. $X_n = x_n$). Initially we will assume that all we know is $\mu \ge 0$, based on consistency of the interval definition. What is the ML estimator in this case? For the general case of N realizations $x_1, \ldots, x_N$ of the mutually independent $X_1, \ldots, X_N$ (stacked for notational convenience

in the random vectors $\mathbf{x}_N$ and $\mathbf{X}_N$, respectively), one finds for the likelihood L of the event $\mathbf{X}_N = \mathbf{x}_N$ (given a value for μ):

$$L(\mathbf{X}_N = \mathbf{x}_N \; \mu) = f_{\mathbf{X}_N|\mu}(\mathbf{x}_N) = \prod_{n=1}^{N} f_{X_n|\mu}(x_n) = 1/(2\mu)^N \quad \text{if } |x_n| \le \mu,$$

for all $n \le N$ and 0 otherwise. This expression is easily seen to be maximized at $\mu = \max(|x_1|, |x_2|, \ldots, |x_N|$ since if we reduce μ, L becomes 0, whereas if we increase μ, L decreases. Thus the ML estimator, using N realizations, is (see also DeGroot, 1970):

$$\hat{\mu}_{ML}(N) = \arg\max_{\mu > 0} L(\mathbf{X}_N \mid \mu) = \max(|X_1|, |X_2|, \ldots, |X_N|). \tag{28}$$

This is clear since no sample can have larger absolute value than μ so the largest is closest to μ. Now consider the case in which prior knowledge is available such as that μ is above some threshold $\mu_{\min}$, i.e. $\mu \ge \mu_{\min} \ge 0$. We do not rule out the possibility that the statement $\mu \ge \mu_{\min}$ is false; the consequences of incorrect knowledge will be discussed too. Note that the standard case corresponds to $\mu_{\min} = 0$. The Constrained Maximum Likelihood (CML) estimator is:

$$\hat{\mu}_{CML}(N) := \arg\max_{\mu \ge \mu_{\min}} L(\mathbf{X}_N \mid \mu) = \max(\mu_{\min}, |X_1|, |X_2|, \ldots, |X_N|). \tag{29}$$

The validity of this result can be seen as follows. Since $\mu \ge \mu_{\min}$, it follows that $\hat{\mu}_{CML} \ge \mu_{\min}$ w.p. 1. In other words, the prior knowledge can be viewed as an additional realization, say at time 0, stating $X_0 = \mu_{\min}$. On the other hand, if $\hat{\mu}_{ML}(N) > \mu_{\min}$ it follows from the definition of the ML and CML estimator that $\mu_{CML}(N) = \hat{\mu}_{ML}(N)$. This combination yields (29).

So far the (C)ML estimator has been analysed. It can be easily checked that, in this example, MAP/*plu* (with continuous prior behaving linearly for $0 \le \mu \le \mu_{\min} = \theta^0$ and being constant elsewhere) coincides with ML. In addition, we can relate the Constrained Minimum Variance (CMV) Estimator (24) to CML. Introducing the abbreviation $M := \max(\mu_{\min}, |x_1|, |x_2|, \ldots, |x_N|)$, one can write:

$$\mathscr{E}[\mu \mid \mathbf{X}_N = \mathbf{x}_N; \mu \ge \mu_{\min}] = \int_M^\infty \frac{\mu d\mu}{(2\mu)^N} \Big/ \int_M^\infty \frac{d\mu}{(2\mu)^N} = \frac{\mu^{-N+2}/(-N+2)|_M^\infty}{\mu^{-N+1}/(-N+1)|_M^\infty}$$

$$= \frac{N-1}{N-2} \times \frac{M^{2-N}}{M^{1-N}} = \frac{N-1}{N-2} \times \max(\mu_{\min}, |x_1|, |x_2|, \ldots, |x_N|), \quad (N = 3, 4, 5, \ldots),$$

yielding, for $N \ge 3$

$$\hat{\mu}_{CMV}(N) = \mathscr{E}[\mu \mid \mathbf{X}_N; \mu \ge \mu_{\min}] = \frac{N-1}{N-2} \times \hat{\mu}_{CML}(N). \tag{30}$$

This shows that the CMV estimator is less cautious. In the following, the mean and variance of both estimators will be calculated to gain further insight into their relative merits. Since (30) states that for fixed N, CMV is proportional to CML, (29) is the basis of the following calculation of the Lth moment ($L = 1, 2, 3, \ldots$) of $\hat{\mu}_{CML}(N)$:

$$\mathscr{E}[\hat{\mu}_{CML}(N)]^L = \left(\frac{\mu_{\min}}{\mu}\right)^N \times \mu_{\min}^L + \int_{\mu_{\min}}^{\mu} x^L \times d\Pr(|X_1| \le x, |X_2| \le x, \ldots, |X_N| \le x)$$

$$= \left(\frac{\mu_{\min}}{\mu}\right)^N \times \mu_{\min}^L + \int_{\mu_{\min}}^{\mu} x^L \times d[(x/\mu)^N]$$

$$= \frac{N + L \times \left(\frac{\mu_{\min}}{\mu}\right)^{N+L}}{(N+L)} \times \mu^L \tag{31}$$

Note that the first and the second term are the unconstrained and the correction part with respect to $\mu \ge \mu_{\min}$. From (31) it is seen that the relative expected bias β_{CML} of *CML* is

$$\beta_{CML}(N) := (\mathscr{E}\hat{\mu}_{CML}(N) - \mu)/\mu = \frac{\left(\frac{\mu_{\min}}{\mu}\right)^{N+1} - 1}{N+1}$$

This relation confirms that if $\mu \ge \mu_{\min}$ (correct information), $\mathscr{E}\hat{\mu}_{CML}(N)$ converges linearly (at first-order rate) to μ for large N. Note that the expected bias is negative. In addition, using (30) it is seen that the relative expected bias of CMV is

$$\beta_{CMV}(N) := (\mathscr{E}\hat{\mu}_{CMV}(N) - \mu)/\mu = \frac{(N-1)\left(\frac{\mu_{\min}}{\mu}\right)^{N+1} + 2}{(N+1)(N-2)}$$

showing that if $\mu \ge \mu_{\min}$, $\mathscr{E}\hat{\mu}_{CMV}(N)$ converges quadratically to μ for large N. In the CMV case, the expected bias is positive but vanishes more

TABLE 1 RELATIVE EXPECTED BIAS (IN %) AS A FUNCTION OF N FOR ML, CML AND CMV

Number of samples (N)	$\beta_{ML}(N)$ $\mu_{min}=0$	$\beta_{CML}(N)$ $\mu_{min}=0.9\mu$	$\beta_{CMV}(N)$ $\mu_{min}=0$	$\beta_{CMV}(N)$ $\mu_{min}=0.9\mu$
10	−9.1%	−6.2%	2.3%	5.5%
20	−4.8%	−4.2%	0.5%	1.1%
30	−3.2%	−2.0%	0.2%	0.4%

quickly. Both effects are tabulated below for the cases $\mu_{min}=0$ and $\mu_{min}=0.9\mu$.

Table 1 clearly shows that the CMV estimator is more biased for tighter prior knowledge; in the case of the CML estimator it is the other way around. Furthermore, using (31), the variance of $\mu_{CML}(N)$ is

$$\operatorname{var} \hat{\mu}_{CML}(N)$$

$$= \mathscr{E}[\hat{\mu}_{CML}(N)]^2 - [\mathscr{E}\hat{\mu}_{CML}(N)]^2$$

$$\frac{N+2\left(\frac{\mu_{min}}{\mu}\right)^{N+2}}{N+2} - \left[\frac{N+\left(\frac{\mu_{min}}{\mu}\right)^{N+1}}{N+1}\right.$$

$$\frac{N}{(N+1)(N-2)} - \left(\frac{\mu_{min}}{\mu}\right)^{N+1}$$

$$\times \left[\frac{2N+\left(\frac{\mu_{min}}{\mu}\right)^{N+1}}{(N+1)^2} - \frac{2\frac{\mu_{min}}{\mu}}{N+2}\right].$$

Again, the first and the second term are the unconstrained and the correction part of the variance. Since the correction term can be shown to be strictly positive and increasing with μ_{min} (for all $N \geq 3$) if $\mu \geq \mu_{min} > 0$, the CML estimator accuracy increases as a result of tighter knowledge. This is visualized in Fig. 9 for $\mu_{min}=0$ (no knowledge), $\mu_{min}=3\mu/4$ and $\mu_{min}=9\mu/10$, respectively. Note that bias as well as variance reduce with tighter μ_{min}. Furthermore it clearly shows that the impact of prior knowledge is most prominent for *small* sample lengths; asymptotically the "profit" vanishes. The above observations also apply to the CMV estimator because of (30). Since $\hat{\mu}_{CMV}(N)$ is less cautious, its variance is larger, i.e.

$$\operatorname{var} \hat{\mu}_{CMV}(N) = \left(\frac{N-1}{N-2}\right)^2 \times \operatorname{var} \hat{\mu}_{CML}(N).$$

If the prior knowledge is *correct* both constrained estimators inherit the asymptotical unbiasedness of the unconstrained ML estimator:

$$\lim_{N\to\infty} \hat{\mu}_{CMV}(N) = \lim_{N\to\infty} \hat{\mu}_{CML}(N)$$

$$= \lim_{N\to\infty} \hat{\mu}_{ML}(N) = \mu,$$

as can be seen from (29) and (30) together with $\Pr(|X_N| > \mu - \delta) > 1 - \epsilon$ for any $\delta, \epsilon > 0$ if N is sufficiently large. In case of *false information* (i.e. $\mu_{min} > \mu$), it follows from (29) and (30) that the CML as well as the CMV estimator will be biased for all N, i.e. $\hat{\mu}_{CML}(N) = \mu_{min} > \mu$, cf. Fig.

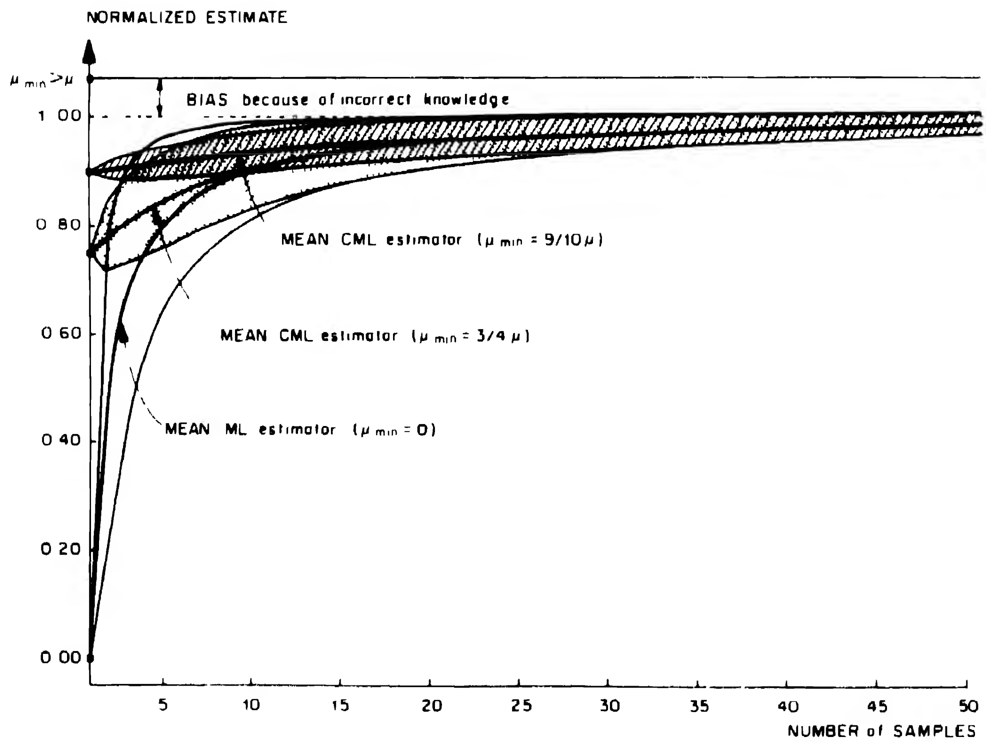


FIG. 9. Mean and variances of ML and CML estimator of μ as a function of N

9. In the CMV case the bias is even bigger; asymptotically the CML level will be met.

To avoid such inconsistencies, one should only use *undeniable* prior knowledge. In this simple example this could easily be detected but the observation applies to the general case too.

5.5. *Some properties of the constrained estimators*

It can be shown that asymptotically (when the number of data items $N \to \infty$) CML, CMV and MAP/*plu* estimators converge with probability one to the conventional ML estimator provided the 'process' is in the admissible model class implying that the true parameter $\theta^* \in \Omega$ That is, the constrained estimators are asymptotically unbiased and efficient when the *a priori* knowledge is correct. This is a direct consequence of the fact that the likelihood will asymptotically become Gaussian distributed, centered at θ^*, with vanishing covariance matrix (e.g. Ljung and Söderström, 1983).

On the other hand, one should be careful when imposing unreliable knowledge since $\theta^* \notin \Omega$ will generate an asymptotic bias, as is illustrated in the previous section. Furthermore, the mean squared error of the constrained estimators is less than that of the ML estimator. Moreover, this advantage can expect to be greater for shorter experiments and when non-stationarities or modelling errors play a dominant role. These observations are illustrated below.

5.6. *Simulation results*

We will briefly discuss a simulation result here. More extensive results are in van Aken (1987). We consider a scalar, first-order ARX process

$$\begin{aligned} y_k = a_p y_{k-1} + b_p u_{k-1} + \xi_k,& \\ a_p = 9/10,& \qquad (32)\\ b_p = 1/10,& \end{aligned}$$

where $\{u_k\}$ is a scalar "Generalized Binary Noise" test signal with unit amplitude and non-switching probability $p := 8/10$. Such a signal can at each time kT "decide" to remain at the actual signal level ($u_k = u_{k-1}$) or switch to an opposite level ($u_k = -u_{k-1}$) with probabilities

$$\begin{aligned} \Pr(u_k = u_{k-1}) &= p, \\ \Pr(u_k = -u_{k-1}) &= 1 - p, \\ 0 < p &< 1. \end{aligned}$$

Generalized Binary Noise has been shown (Tulleken, 1990) to be superior to ordinary Binary Noise (PRBS). In particular, it is asymptotically (i.e. for large experiments) *optimal* for stable first-order ARMAX processes when p equals $(1 + a_p)/2$. Since the most relevant simulation experiment length is rather small, the non-switching probability is selected somewhat less than $(1 + 9/10)/2 = 0.95$. The noise ξ is chosen white Gaussian $\xi \sim \mathcal{N}(0, \frac{1}{2})$ of which the outcomes are not measurable The variance is chosen such that the signal-to-noise ratio is about unity (as much induced signal as noise on output) which is often the case under realistic conditions in the process industry. Candidate models of similar ARX format are considered:

$$y_k = a y_{k-1} + b u_{k-1} + \xi_k, \quad (a, b \in \mathbb{R}). \quad (33)$$

The unknown a and b have to be estimated. In order to define a GB model class, some crucial knowledge is imposed. In this example we consider knowledge of *open-loop stability and bounds on the stationary gain* $\mathcal{G} := b/(1 - a)$:

$$1/5 \le \mathcal{G} \le 5.$$

This results (see (9) and (11)) in a Grey-Box model set Ω, given by:

$$\begin{aligned} \Omega := \{a, b \in \mathbb{R} \mid -1 \le a \le 1,& \\ (1 - a) \le 5b \le 25(1 - a)\}.& \end{aligned}$$

Note that the process is indeed stable and has unity stationary gain, so the true process parameter is in Ω. Now, using 100 Monte-Carlo simulation trials in order to scan the dependency on the particular data sets realized, the stochastic properties of ML (which of course equals *LS* in this case), CML and CMV for this particular process are obtained with good approximation. The expected values of the parameter estimators $\hat{a}$, $\hat{b}$ and $\hat{\mathcal{G}} := \hat{b}/(1 - \hat{a})$ for ML, CML and CMV are shown in Fig. 10(a)–(c) (solid lines). In addition, the successive approximated 95% probability curves are shown using dashed lines. The results are represented as a function of the experiment length, expressed as a multiple of the 95% process settling time τ^{95} ($\tau^{95} \approx [-3/\ln 0.9]T \approx 28T$). For the process industry, the most interesting situation is around indices 2, 3 and 4. It can be seen that all three estimators asymptotically converge to the correct values, indicated with dotted lines. However, for short experiments the constrained estimators and in particular CMV are more biased than ML. On the other hand, the estimator uncertainty (in practical terms, the realization dependency) is drastically reduced for CML and especially for CMV. This improvement is most significant in the estimators of b and $\mathcal{G}$, since low-frequency characteristics are much better modelled. Note

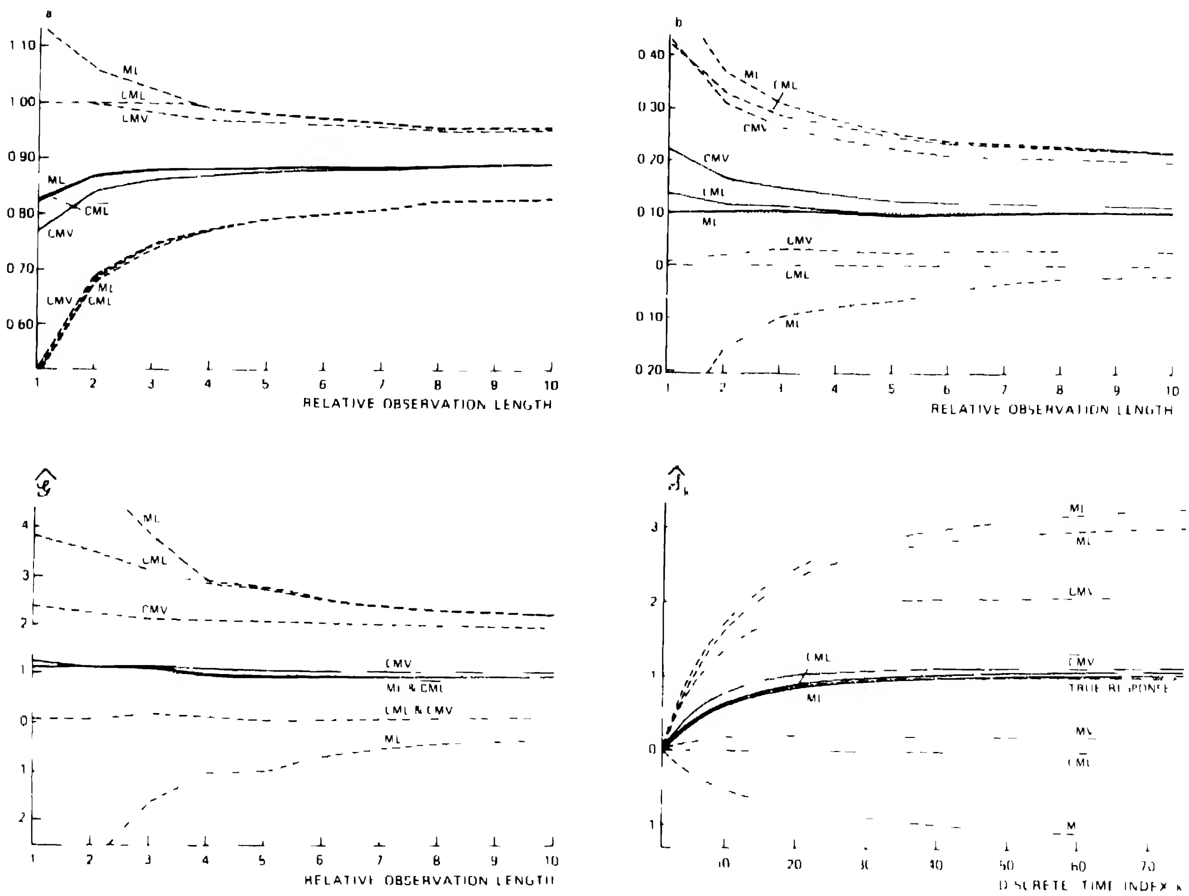


FIG. 10. ML, CML and CMV estimators for a, b, gain $\mathscr{G}$ and $\mathscr{S}_k$, respectively.

that ML for short experiments indeed suffers from a high probability of selecting non-sense models being unstable or having wrong sign of static gain. For step response and frequency response estimators similar observations can be made. Figure 10(d) presents the distribution of ML, CML and CMV step response parameter estimators for an observation length $4\tau^{95}$. These results suggest that, on average, CMV is much more reliable than ML, which can be evidenced with theoretical arguments and more complicated simulations as well. Obviously, the increased model accuracy has important consequences for the success and robustness of controller design.

5.7. *Impact of Grey-Box approach on controller design*

Consider the first-order process (32) (with *true* parameters $a_p = 0.9$ and $b_p = 0.1$) and the corresponding model (33) (using estimators $\hat{a}$ and $\hat{b}$, taking values $\bar{a}$ and $\bar{b}$). On the basis of this identified model we want to design a simple fixed, proportional controller law for (32):

$$u_k := -\mathscr{G}y_k. \tag{34}$$

It is sufficient for our goal to select the gain $\mathscr{G}$ using conventional pole-placement, depending on the identified parameters $\bar{a}$ and $\bar{b}$, i.e. $\mathscr{G} = \mathscr{G}(\bar{a}, \bar{b})$. For notational convenience we skip the arguments if no confusion can arise. The *model* (33) combined with controller (34) will yield a closed-loop first-order system having its pole $\bar{a} - \mathscr{G}\bar{b}$ at some desired location λ, if $\mathscr{G} := (\bar{a} - \lambda)/\bar{b}$. Note the extremes of the open-loop ($\mathscr{G} = 0$; $\lambda = a$) and the minimum-variance (dead-beat) case ($\mathscr{G} = \bar{a}/\bar{b}$; $\lambda = 0$). *If the model were perfect*, the aim would be realized. However, in reality the modelling error will introduce a more or less deteriorated controller performance. The installation of such a controller (34) in combination with the *process* (32) does not even guarantee closed-loop stability, since the magnitude of the *true* closed-loop pole $a_p - \mathscr{G}(\bar{a}, \bar{b})b_p$ may be greater than one. Modern, attractive techniques for robust (e.g. $\mathscr{H}_\infty$) controller design have become available (e.g. Kwakernaak, 1986; Doyle *et al.*, 1989; Djavdan *et al.*, 1989) which account for the effects of uncertainty in model, process or performance requirements. However, these methods inevitably lead to conservative and thus inefficient control performance if the uncertainty is relatively large. Therefore the efforts to

improve upon model identification techniques remain essential. Below it is illustrated that Grey-Box modelling can have considerable benefits in terms of improved controller design.

Since $a_p = 0.9$ and $b_p = 0.1$, only proportional controllers with gain $\mathcal{G}$ satisfying

$$-(1 - a_p)/b_p < \mathcal{G} < (1 + a_p)/b_p,$$

i.e. $-1 < \mathcal{G} < 19$, will stabilize the process. In addition, one could ask: when are we performing better than open-loop, i.e. when do we decrease the asymptotic output variance $\lim_{k \to \infty} \mathcal{E} y_k^2$ with the selected controller? Since, using (32) and (34),

$$\mathcal{E} y_k^2 = \mathcal{E}[(a_p - \mathcal{G} b_p)^2 y_{k-1} + \xi_k]^2,$$

yielding

$$\lim_{k \to \infty} \mathcal{E} y_k^2 = \sigma^2 / [1 - (a_p - \mathcal{G} b_p)^2],$$

the variance decreases if $1 - (a_p - \mathcal{G} b_p)^2 > 1 - a_p^2$, i.e. $0 < \mathcal{G}(\bar{a}, \bar{b}) < 2a_p/b_p = 18$. Note that the *minimum-variance* gain $a_p/b_p = 9$ is in the middle of this interval. Finally, it is relevant to see for which sets of differently estimated parameters (say a and b) the design generates the same gain as when $\bar{a}$ and $\bar{b}$ were being used. This is the case when $\mathcal{G}(a, b) = (a - \lambda)/b = \mathcal{G}(\bar{a}, \bar{b})$, yielding equations of straight lines $b = (a - \lambda)/\mathcal{G}(\bar{a}, \bar{b})$, which all pass through $(\lambda, 0)$. The straight lines shown in Fig. 11 for the minimum-variance design case ($\lambda = 0$) refer to the parameters sets yielding closed-loop marginal stability ($\mathcal{G} = -1$ and $\mathcal{G} = 19$), no variance reduction ($\mathcal{G} = 0$ and $\mathcal{G} = 18$, "break-even") and minimum-variance ($\mathcal{G} = 9$). In addition, the approximated means and 95%-equi-density contours of the ML and CMV estimator of $[\hat{a}, \hat{b}]'$ are shown. These results were gained using Monte-Carlo simulation where the conditions of the previous section applied. Specifically, an experiment length of three times the 95% settling time (85 samples) was selected with a unity signal-to-noise ratio. In the CMV case the Grey-Box model set $\Omega := \{a, b \in \mathbb{R} \mid -1 \le a \le 1; (1 - a) \le 5b \le 25(1 - a)\}$, of which a corner is visible in Fig. 11, was used again. Figure 11 shows the effect of model uncertainty on the closed-loop performance. It can be clearly seen that the probability of not ending up with an unsuccessful or even instable closed-loop process (which equals the integral of the density over the region between the "break-even" $\mathcal{G} = 0$ and $\mathcal{G} = 18$ lines) is considerably lower in the CMV case, even when ML-identified models showing negative sign of static gain would be skipped in advance (which are situated below the horizontal axis). This effect is even more pronounced for $\lambda > 0$. In that case the lines (of model parameters generating equivalent closed-loop behaviour) uniformly shift a distance λ to the right, covering an increasing part of the CMV density.

5.8. *Application of CML to the distillation column field data*

After the above discussion of theoretical and simulation results of Grey-Box estimators, it is interesting to see how these methods improve the identification of a real process. We will focus on the distillation column and associated input-output data which was described in Figs 1 and 2. Specifically, the dynamics from set point of the reflux flow controller ϕ_R to the top temperature T_{34} will be investigated; additional

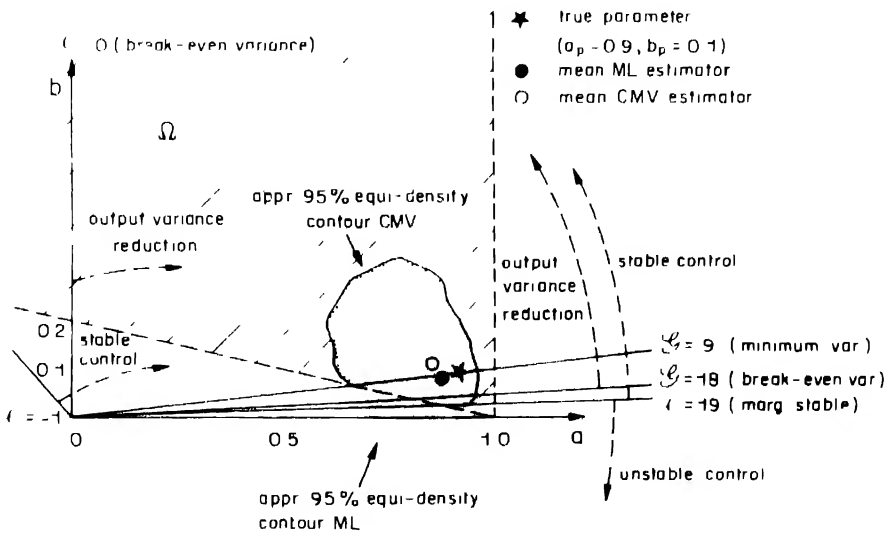


FIG. 11 The impact of Grey-Box results on the quality of controller design

examples are in Peeters (1987). A scalar ARMAX(2, 2, 2) model was postulated, i.e.

$$y_k + a_1 y_{k-1} + a_2 y_{k-2} = b_1 u_{k-1} + b_2 u_{k-2} + \xi_k + c_1 \xi_{k-1} + c_2 \xi_{k-2}.$$

It was known that the process dynamics were open-loop *stable* having *negative* stationary gain $\mathcal{G}$ (since more reflux cools the top of the column), which was certainly above $-\frac{1}{2}$. Using (10), or Example 3 and (11) for $n = 2$, together with the fact that only sound discrete-time poles were acceptable (implying $a_1 \le 0$ and $a_2 \ge 0$, cf. Theorem 2) the admissible parameter set Ω was given by

$$\Omega := \left\{ \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ c_1 \\ c_2 \end{bmatrix} \in \mathbb{R}^6 \quad \begin{matrix} a_1 \le 0 \\ 0 \le a_2 \le 1 \\ 1 \pm a_1 + a_2 \ge 0 \\ -(1 + a_1 + a_2)/2 \le b_1 + b_2 \le 0 \end{matrix} \right\}.$$

Over the full experiment length of 1200 samples the transients of three estimators were compared: ML, CML and CML$^{\text{rec}}$. The last two estimators use the prior knowledge that the parameters must be in Ω. Here CML is the "batch" version where for each extended experiment length the likelihood is combined with a prior uniformly distributed on Ω, yielding a posterior of which the maximization is giving rise to either ML or projected ML results, cf. (21). CML$^{\text{rec}}$ does the same, but uses possibly projected, intermediate results for subsequent prior definition in the recursive updating. The reader may verify that CML$^{\text{rec}}$ is the MAP/*rg* estimator which maximizes, for each experiment length, the posterior associated with the restricted Gaussian prior (19), where θ^0 is the last estimate and Π the associated variance. In Figs 12(a) and (b) the three different estimators $\hat{a}_1$ and $\hat{a}_2$ of a_1 and a_2 are shown, whereas Fig. 12(c) presents the estimated stationary gain $\hat{\mathcal{G}} := (\hat{b}_1 + \hat{b}_2)/(1 + \hat{a}_1 + \hat{a}_2)$. Figure 12(d) indicates that the CML$^{\text{rec}}$ estimator had to perform a huge number of projections towards the edge of Ω (FLAG = 1) during the recursive identification process because of detected model inconsistencies. Some of them are apparent from Fig. 12, e.g. most of the initial 100 ML estimates of $\mathcal{G}$ were (way) out of bounds. A more serious, persistent deficiency is shown in Fig. 12(b), since the ML estimate of $-a_2$ is strictly positive. This reflects a physically pathological situation, with one of the two estimated poles being negative. Since the Grey-Box set Ω excludes such models, CML had to correct always and CML$^{\text{rec}}$ often (Fig. 12(d)) for that tendency In the CML case this results in this simple example in subsequent

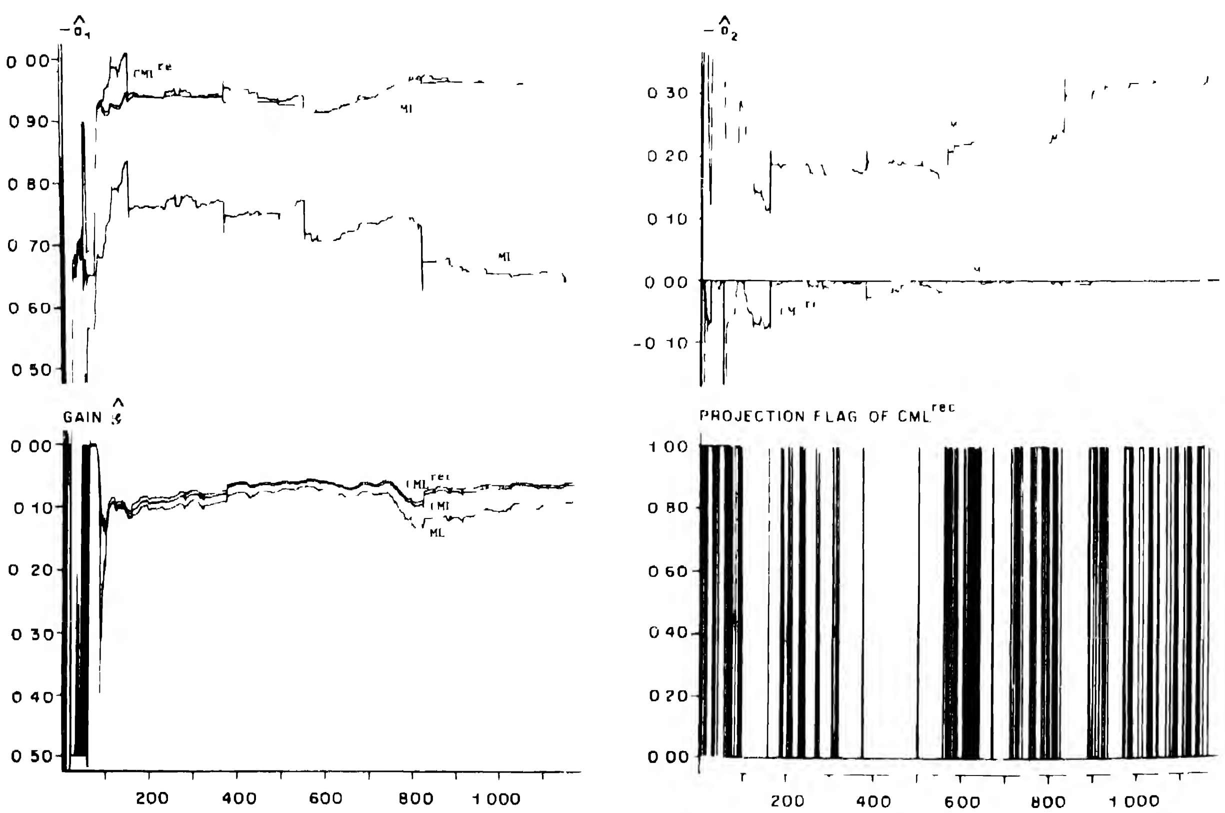


FIG. 12. ML, CML and CML$^{\text{rec}}$ estimators for reflux to top temperature model

projection onto the relevant constraint ($a_2 \geq 0$), whereas CML^{rec} is penetrating the acceptable region, with a tendency towards this edge. Effectively this implies that in this case no second-order *AR* dynamics could be properly fitted because of overparametrization. As can be seen from the CML estimates of $-a_1$ (which equals the discrete-time pole of the resulting *first-order* model) the physical relevance of the AR dynamics is enhanced considerably: the associated time constant of the underlying continuous-time model is about 20 samples instead of about two in the ML case.

Amongst other things, this example clearly shows how model reduction of too complicated models can be realized in a natural fashion by exploiting the physical information. Additional experiments (see also Peeters (1987)) have confirmed that the Grey Box approach is capable of enhancing a number of characteristics of the model in practical cases considerably. Essential features such as stability, sign of stationary gains and occurrence of sound poles are guaranteed with improved rate of convergence.

6. CONCLUSIONS

In this paper, Grey Box statistical estimation is discussed where a linearly parametrized dynamic regression model is identified which explains the data best but is also consistent with specified physical knowledge. It is shown that important physical knowledge such as stability (or even an admissible area for pole location), as well as signs or bounds on steady gains can be translated into inequalities that are linear in the model parameters. Using Bayesian inference, Constrained Maximum and Average *A Posteriori* estimators are introduced. Explicit solutions for the special case of Gaussian likelihood and uniform, as well as *plu*riform (piece-wise linear with uniform component) prior are presented, resulting in a CML, CMV as well as MAP/*plu* estimator with promising characteristics. Theoretical considerations, a simulation result and a practical application to a distillation process show the advantage of the constrained estimators under non-ideal, realistic experiment conditions. Considerable variance reductions at the cost of a somewhat larger bias can be achieved, indicating the potential for advanced control applications where the availability of an accurate model is essential. It may imply that, e.g. an adaptive, automated design will come more into reach.

Acknowledgement—The author likes to thank Prof. A. Bagchi (Twente Univ.), P. A. van Aken and R. I. M. Peeters (both former M.Sc. students under guidance of the author), J. Linders (Shell Internationale Petroleum Maatschappij) and M. P. Voetter (K.S.L.A.) for the collaboration and helpful discussions. He is specially indebted to Prof. G. J. Olsder and Prof. H. B. Verbruggen (both Delft Univ. of Techn.) as well as Prof. J. Grievink (Delft Univ. of Techn., formerly K.S.L.A.) for their support, advice and review of the manuscript.

Finally, the anonymous reviewers are gratefully acknowledged for their constructive criticism and helpful suggestions.

REFERENCES

Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans.* **AC-19**, 716.

Akaike, H. (1978). A new look at the Bayes procedure. *Biometrica* **65**, 53–59.

Akaike, H. (1986). The selection of smoothness priors for lag estimation. In P. Goel and A. Zellner (Eds) *Bayesian Inference and Decision Techniques*, pp. 109–118. Elsevier, Amsterdam.

Aken, P. A. van (1987). Constrained identification using physical knowledge. Masters thesis, Dept. of Mathematics, Twente Univ. of Technology.

Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press, New York.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley, New York.

Casti, J. L. (1977). *Dynamical Systems and their Applications*. Academic Press, New York.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw Hill, New York.

Djavdan, P., H. J. A. F. Tulleken, M. H. Voetter, H. B. Verbruggen and G. J. Olsder (1989). Probabilistic robust controller design. *28th CDC Tampa*, pp. 2164–2174.

Doyle, J. C., K. Glover, P. M. Khargonekar and B. A. Francis (1989). State space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Trans.* **AC-34**, 831.

Eykhoff, P. (1974). *System Identification, Parameter and State Estimation*. Wiley, London.

Franklin, G. F. and J. D. Powell (1980). *Digital Control of Dynamic Systems*. Addison Wesley, New York.

Gertler, J. (1979). A constrained minimum variance input–output estimator for linear dynamic systems. *Automatica* **15**, 353–358.

Goodwin, G. C. and R. L. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Math. in Sc. & Engineering, Vol. 136. Academic Press, New York.

Goodwin, G. C. and K. S. Sin (1984). *Adaptive Filtering, Prediction and Control*. Inf. & Systems Sc. Series, Prentice Hall, Englewood Cliffs, NJ.

Goodwin, G. C. and Salgado (1989). A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *Int. J. Adaptive Control and Signal Proc.* **3**, 333–356.

Jury, E. I. (1974). *Inners and Stability of Dynamical Systems*. Wiley, New York.

Jørgensen, S. B., L. Goldschmidt and I. Hallager (1985). Sparse process modelling for robust adapative surveillance and control of a binary distillation column. *IFAC Adaptive Control of Chemical Process*, p. 184, Frankfurt.

Kwakernaak, H. (1986). A polynomial approach to minimax frequency domain optimization of multivariable feedback systems. *Int. J. Contr.* **44**, 117–156.

Liew, C. K. (1976). Inequality constrained least squares estimation. *J. Am. Stat. Association* **71**, 746–751.

Ljung, L. and T. Soderstrom (1983). *Theory and Practice of Recursive Identification*. MIT, Cambridge.

Nemir, D. C. and R. L. Kashyap (1986). Certainty equivalence and the minimax principle in self tuning control. *IEEE Trans. AC* **6**, 587.

Peeters, R. I. M. (1987). Constrained Le. squares estimation of ARLMAX models. Masters thesis, Dept. of Mathematics, Delft Univ. of Techn.

Peterka, V. (1981) Bayesian approach to system identification. In P. Eykhoff (Ed.), *Trends and Progress in System Ident.*, pp. 239-304. Pergamon, Oxford.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models and Applications.* John Wiley, New York.

Shieh, L. S. and S. Sacheti (1978). A matrix in the block Schwarz form and the stability of matrix polynomials. *Int. J. Control,* **25,** 245–259.

Shieh, L. S., M. M. Mehio and H. M. Dib (1987). Stability of the second-order matrix polynomial. *IEEE Trans AC,* **32,** 231–233.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice Hall, London.

Tulleken, H. J. A. F. (1987). A multivariable adaptive forgetting strategy for on-line identification in a self-tuning controller. *J. A,* **28** (4).

Tulleken, H. J. A. F. (1989). Grey-Box Mod. & Identification using physical knowledge and Bayesian techniques. *IFAC Symp. on Adaptive Control and Signal Processing,* Glasgow, pp. 591–598.

Tulleken, H. J. A. F. (1990). Generalized binary noise test-signal concept for improved identification-experiment design. *Automatica,* **26,** 37–49.

Tulleken, H. J. A. F. (1992). Grey-box Modelling and Identification Topics, Doctoral Thesis, Delft University of Technology.

Wahlberg, B. and L. Ljung, (1986). Design variables for bias distribution in transfer function estimation. *IEEE Trans. AC,* **2,** 134.

Automatica, Vol 29, No 2 pp 309–322, 1993
Printed in Great Britain

0005 1098/93 $6 00 + 0 00

An Integrated Collision Prediction and Avoidance Scheme for Mobile Robots in Non-stationary Environments*

K. J. KYRIAKOPOULOS† and G N. SARIDIS‡

A scheme integrating simple and fast collision prediction and a reactive type (feedback solution) of avoidance for mobile robots in environments containing moving obstacles is computationally efficient and appropriate for real time implementation.

Key Words—Mobile robots, motion planning, collision prediction, collision avoidance, potential functions, Lyapunov stability

Abstract —A formulation that makes possible the integration of collision prediction and avoidance stages for mobile robots moving in general terrains containing moving obstacles is presented A dynamic model of the mobile robot and the dynamic constraints are derived Collision avoidance is guaranteed if the distance between the robot and a moving obstacle is nonzero A nominal trajectory is assumed to be known from off-line planning The main idea is to change the velocity along the nominal trajectory so that collisions are avoided A feedback control is developed and local asymptotic stability is proved if the velocity of the moving obstacle is bounded Furthermore, a solution to the problem of inverse dynamics for the mobile robot is given Simulation results verify the value of the proposed strategy

1 INTRODUCTION

THE PROBLEM OF efficiently planning the motion of a mobile robot in an environment containing moving obstacles (Fig. 1) is difficult and computationally intensive. Although it has been stated as early as 1984 (Freund and Hoyer, 1984) *ad hoc* solutions were given (Liu *et al.* (1989); Tournassoud (1986)). Reif and Sharir (1985) gave an algorithmic solution to the problem but they were restricted to some categories of shapes of objects and their approach is not suitable for an on-line implementation. Search based approaches for solving the above problem have also been presented in Fujimura and Samet (1990) and Shih *et al.* (1990). On the other hand, Kant and Zucker (1984, 1986, 1988) used the decomposition of the motion planning problem to the find-path, and move-along-path problems. Their proposal is that the avoidance of moving obstacles can be done by adjusting the speed along the geometric path. The same approach was adopted in Wu and Jou (1988) and in Griswold and Eem (1990). The basic idea of this approach is utilized in this work.

To facilitate a fast solution a hierarchical decomposition has been proposed Kant and Zucker (1988) and adopted and extended in our work. The problem is divided as follows.

- Off-line path and motion planning.
- On-line motion replanning.

In the off-line stage, two problems are solved; First, path planning, the "find path" problem, i.e. the search for a connected curve $r(s) = [x(s)y(s)z(s)]^T$ on the terrain described by a surface $g(x, y, z) = 0$; s is the trajectory parametrization variable (e.g. path length). The path should connect the initial and target points without colliding with the stationary objects while satisfying certain criteria. Second, motion planning, that is to find a "nominal" motion function $s_n(t)$ along this path that does not violate the kinematic and dynamic constraints of the robot, and some performance criterion (e.g. time) is minimized.

The subject of this paper is the development of an algorithm for the on-line stage. This algorithm has to act in a supervisory mode during motion execution and make sure that the robot is going to move from its current state (position s_0, velocity v_0) to the target one

* Received 1 October 1991, revised 11 April 1992; received in final form 9 June 1992 The original version of this paper was not presented at any IFAC meeting This paper was recommended for publication in revised form by Editor A P Sage.

† New York State Center for Advanced Technology in Automation and Robotics, Rensselaer Polytechnic Institute Troy, NY 12180-3590, U S.A.

‡ NASA Center of Intelligent Robotic Systems for Space Exploration, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, U S.A

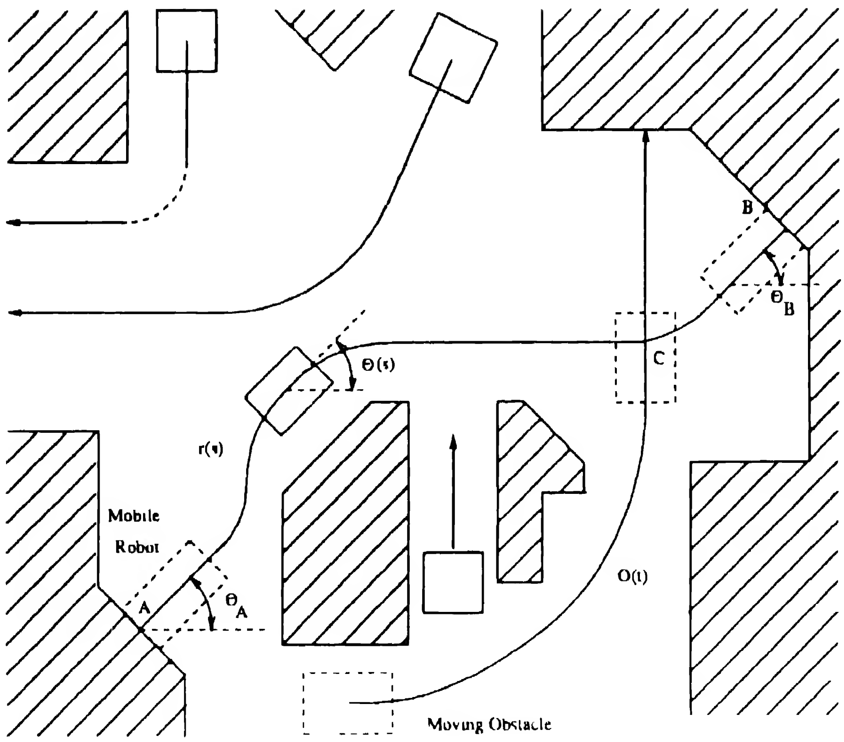


FIG. 1. Environment with a mobile robot and multiple moving obstacles

(position s_f, velocity v_f) avoiding collisions with those moving obstacles with which collision is predicted based on sensory input. The basic idea is to alter the velocity along the path $r(s)$ without changing its geometry. This is useful only if the following fundamental assumption is satisfied:

Temporary obstruction assumption. The mobile robot moving along path $r(s)$ can only be obstructed during a bounded amount of time, i.e. the moving object is assumed not to permanently stay on, or move parallel to $r(s)$.

The new plan must satisfy the dynamic constraints and stay as close as possible to the nominal plan.

In our prior developments the objects were modeled as convex polyhedra and based on this assumption an approach to predict collisions (Kyriakopoulos and Saridis (1990a, 1992a)) was developed. For the case of a planar terrains a real time collision avoidance scheme, the Minimum Interference Strategy (MIS) and the Optimal Control Strategy (OCS) (Kyriakopoulos and Saridis (1990b, 1991, 1992b)), have been proposed. There, in addition to collision avoidance, time consistency with the nominal plan was sought.

In this paper the potential fields strategy (PFS) is presented. Potential fields were originally introduced by Khatib (1985) and furthermore investigated by Rimon and Koditschek (1989) and applied for mobile robots by Kant and Zucker. It is a computationally inexpensive but local method providing only collision avoidance capabilities while performance is not guaranteed.

Here, a reformulation of the potential fields approach is attempted, so that performance as well as collision avoidance are sought without significantly increasing the computational load. Performance is expressed in terms of proximity of the final time of the new plan to the final time of the off-line plan. Performance is achieved by the following.

- Introducing a simple collision prediction stage in order to determine if action should be taken.
- Utilizing the information of the nominal plan.

In Section 2 the definitions, modeling and the mathematical problem are stated. The theoretical analysis and the presentation of the Potential Fields Strategy (PFS) is done in Section 3. Finally, in Sections 4 and 5 simulation results and suggestions for future research are presented, respectively.

2. PROBLEM FORMULATION

The mobile robot is a kinematic mechanism composed of the body and the rolling wheels. Its kinematics and dynamics can be modeled based on the necessary assumption that the wheels are

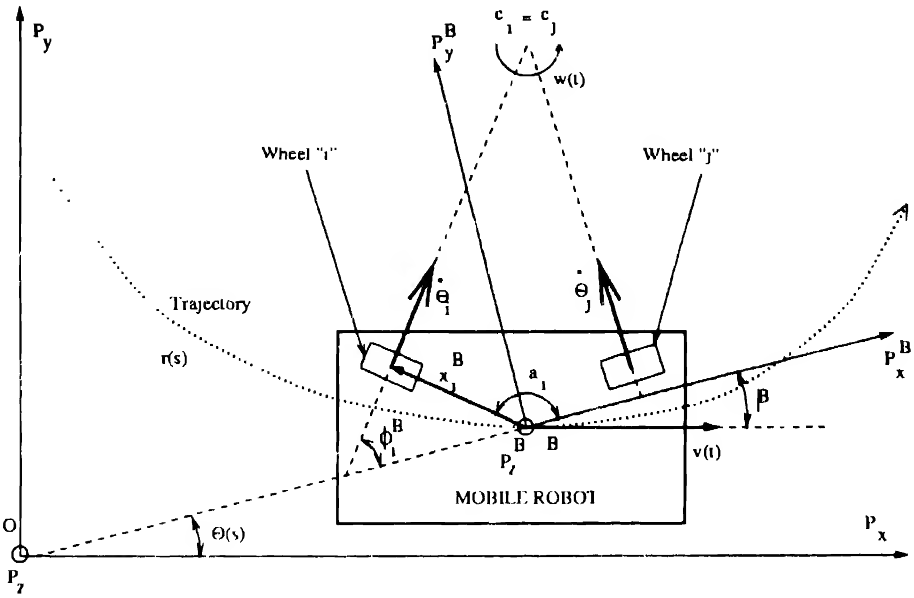


FIG. 2 Top view and kinematic parameters of a mobile robot

ideally rolling The conditions to achieve rolling without slipping and skidding are presented

2.1 *Kinematic modeling*

Although the development of Alexander and Maddocks (1989) is based on the assumption of planar motion their analysis holds for the case of general terrains with the additional assumption that the curvature of the floor is such that sufficient contact of the wheels is guaranteed. Consider the top view of a mobile robot (Fig. 2) moving along a trajectory described in a parametric form by $r(s)$ where s is the parametrization variable. The world coordinate frame (p_x, p_y, p_z) with unit vectors $(\vec{i}, \vec{j}, \vec{k})$ is fixed at O. The body coordinate frame (p^B_x, p^B_y, p^B_z) is attached to the body at point B and moves along $r(s)$.

The motion of the robot is described by $r(s)$ and the orientation angle function $\Theta(s)$, defined as the angle between p_x and p^B_x. The instant translational velocity of the robot with respect to the world coordinate frame is $\vec{v}(t)$ and its rotational velocity is $\vec{\omega}(t) = \omega(t)\vec{k}$. The signed angle between $\vec{v}(t)$ and p^B_x is β.

Two wheels are located at x_i, x_j w.r.t. the world coordinate frame and at x^B_i, x^B_j w.r.t. the body coordinate frame. a_i, a_j are the signed angles between x^B_i, x^B_j and p^B_x, respectively. The angles ϕ^B_i and ϕ^B_j between their axis of rotation and p^B_x, are called steering angles. Vectors $\dot{\theta}_i$ and $\dot{\theta}_j$ represent the angular velocity of the rolling wheels.

The centers c_i and c_j of planar rotation, corresponding to each wheel are:

$$c_i = x_i + \frac{\dot{\theta}_i r}{\omega},$$

$$c_j = x_j + \frac{\dot{\theta}_j r}{\omega},$$

where r is the radius of the wheel. In order to have ideal rolling of the robot, the centers c_i, c_j must be identical, and therefore

$$\omega(x_i - x_j) = r(\dot{\theta}_j - \dot{\theta}_i). \tag{1}$$

This is the rolling compatibility condition that guarantees rolling without sliding. In Appendix A the inverse kinematics are presented based on the development of Alexander and Maddocks (1989).

2.2. *Velocity bounds*

To avoid skidding, the centrifugal forces should not saturate the available friction between the wheels and the floor. In order to satisfy this condition, a bound $v_{skid}(s)$ on the velocity must be set at each point s. A vehicle with n_w wheels moving along path $r(s) = [x(s)y(s)z(s)]^T$† on a terrain described by the surface $g(x, y, z) = 0$ is considered. If the assumption that $g(x, y, z)$ is reasonably smooth is made, the contact points of wheels are planar. A coordinate frame p^0_x, p^0_y, p^0_z is attached to the center of mass G of the vehicle (see Fig. 3). This is essentially the same with the frame p^B_x, p^B_y, p^B_z

† The path will be sometimes denoted as $r(s) = [x(s)y(z)z(s)]^T$ and sometimes $\vec{r}(s) = x(s)\vec{i} + y(s)\vec{j} + z(s)\vec{k}$.

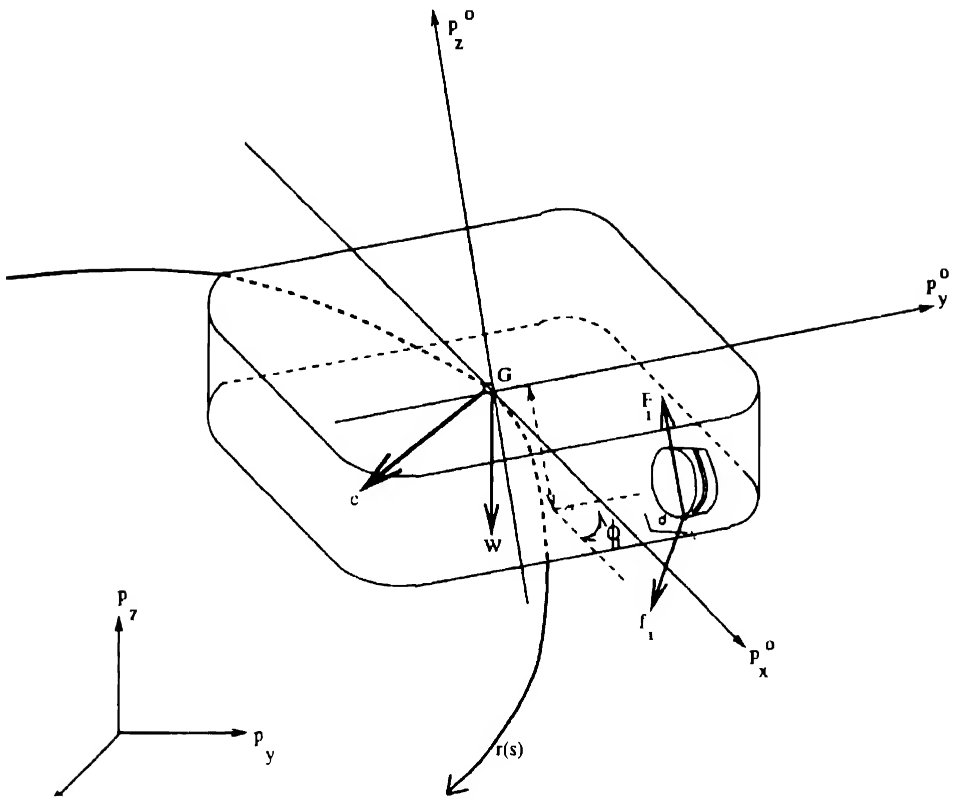


Fig. 3 Forces acting on a mobile robot moving on general terrain

of the previous section assuming $B \equiv G$ and $\beta = 0$. If p_x^0 is tangent to $r(s)$, p_z^0 is normal to the plane tangent to the surface of the floor, and p_y^0 is normal to both then the unit vectors $(\vec{i}^0, \vec{j}^0, \vec{k}^0)$ corresponding to (p_x^0, p_y^0, p_z^0) are

$$\vec{i}^0(s) = \frac{d\vec{r}(s)}{ds}, \quad \vec{j}^0(s) = \vec{k}^0 \times \vec{i}^0, \quad \vec{k}^0(s) = \frac{\vec{\nabla} g}{\|\vec{\nabla} g\|} \tag{2}$$

Since coordinate frame p_x^0, p_y^0, p_z^0 moves w.r.t time

$$\begin{aligned} \frac{d\vec{i}^0}{dt} &= \vec{\omega} \times \vec{i}^0 \Rightarrow \frac{d\vec{i}^0}{ds}\dot{s} = \vec{\omega} \times \vec{i}^0, \\ \frac{d\vec{j}^0}{dt} &= \vec{\omega} \times \vec{j}^0 \Rightarrow \frac{d\vec{j}^0}{ds}\dot{s} = \vec{\omega} \times \vec{j}^0, \\ \frac{d\vec{k}^0}{dt} &= \vec{\omega} \times \vec{k}^0 \Rightarrow \frac{d\vec{k}^0}{ds}\dot{s} = \vec{\omega} \times \vec{k}^0, \end{aligned} \tag{3}$$

where $\vec{\omega}$ is the vector of angular velocity. If equations (3) are solved† then $\vec{\omega}$ has the form

$$\vec{\omega} = \vec{\Omega} \cdot \dot{s}. \tag{4}$$

† This set of equations is not actually overdetermined This is because the $z(s)$ of vector $\vec{r}(s)$ can be determined by $x(s)$, $y(s)$ and $g(x, y, z) = 0$

Similarly,

$$\begin{aligned} \frac{d^2\vec{i}^0}{dt^2} &= \vec{\omega} \times (\vec{\omega} \times \vec{i}^0) + \frac{d\vec{\omega}}{dt} \times \vec{i}^0 \Rightarrow \frac{d\vec{i}^0}{ds}\ddot{s} + \frac{d^2\vec{i}^0}{ds^2}\dot{s}^2 \\ &= \vec{\omega} \times (\vec{\omega} \times \vec{i}^0) + \frac{d\vec{\omega}}{dt} \times \vec{i}^0, \\ \frac{d^2\vec{j}^0}{dt^2} &= \vec{\omega} \times (\vec{\omega} \times \vec{j}^0) + \frac{d\vec{\omega}}{dt} \times \vec{j}^0 \Rightarrow \frac{d\vec{j}^0}{ds}\ddot{s} + \frac{d^2\vec{j}^0}{ds^2}\dot{s}^2 \\ &= \vec{\omega} \times (\vec{\omega} \times \vec{j}^0) + \frac{d\vec{\omega}}{dt} \times \vec{j}^0, \\ \frac{d^2\vec{k}^0}{dt^2} &= \vec{\omega} \times (\vec{\omega} \times \vec{k}^0) + \frac{d\vec{\omega}}{dt} \times \vec{k}^0 \Rightarrow \frac{d\vec{k}^0}{ds}\ddot{s} + \frac{d^2\vec{k}^0}{ds^2}\dot{s}^2 \\ &= \vec{\omega} \times (\vec{\omega} \times \vec{k}^0) + \frac{d\vec{\omega}}{dt} \times \vec{k}^0, \end{aligned} \tag{5}$$

and the angular acceleration has the form

$$\vec{\alpha} = \frac{d\vec{\omega}}{dt} = \vec{A}_2(s)\ddot{s} + \vec{A}_1(s)\dot{s}^2. \tag{6}$$

Unit vector $\vec{c}$ lies on the (p_y^0, p_z^0) plane and points towards the center of curvature or $r(s)$.

$$\frac{d^2\vec{r}/ds^2}{\|d^2\vec{r}/ds^2\|} \tag{7}$$

The acting forces on the vehicle are; the weight $\vec{w} = mg\vec{k}$, the reaction forces $\vec{F}_i = F_i\vec{k}^0$ $i = 1, \ldots, n_w$ from the floor to the wheels and the friction forces $\vec{f}_i = f_{ix}\vec{i}^0 + f_{iy}\vec{j}^0$, $i = 1, \ldots, n_w$ between the floor and the wheels. When the robot moves with a speed $\dot{s}$ and acceleration $\ddot{s}$ a centripetal force

$$\vec{F}_c = m\dot{s}^2 f(s)\vec{c}, \tag{8}$$

an inertial force

$$\vec{F}_1 = m\ddot{s}\vec{i}^0, \tag{9}$$

and an inertial torque of the form

$$\vec{M}_\theta = I\vec{\alpha} + \vec{\omega} \times (I\vec{\omega}) \overset{(4,6)}{=} \vec{M}_2(s)\ddot{s} + \vec{M}_1(s)\dot{s}^2, \tag{10}$$

should be exerted. (I is the inertia matrix of the mobile robot around its center of gravity.)

The motion equations of the vehicle are

$$\sum^{n_w} (\vec{f}_i + \vec{F}_i) + \vec{w} + \vec{F}_c + \vec{F}_1 = 0, \tag{11}$$

$$\sum_{i=1}^{n_w} (\vec{f}_i + \vec{F}_i) \times (\vec{x}_i^0 - l\vec{k}^0) + \vec{M}_\theta = 0, \tag{12}$$

where l is the height difference between the contact points of the wheels and the center of mass G.

The slide constraints for the wheels are:

$$|f_i| = \sqrt{(f_{ix})^2 + (f_{iy})^2} \le \eta F_i, \quad i = 1, \ldots, n_w, \tag{13}$$

$$F_i \ge 0, \quad i = 1, \ldots, n_w, \tag{14}$$

where η is the friction coefficient between the wheels and the floor.

The actuator constraints are

$$-\mathscr{F}_2 \le u_i = f_{ix} \sin \phi_i^0 + f_{iy} \cos \phi_i^0 \le \mathscr{F}_1, \quad i = 1, \ldots, n_w, \tag{15}$$

where $(-\mathscr{F}_2, \mathscr{F}_1)$ are the brake and accelerating bounds of the actuators, respectively.

The maximum velocity so that skidding is avoided is given by the mathematical program

$$v^2_{\text{skid}}(s) = \max \{\dot{s}^2 / (11)\text{–}(15) \text{ are satisfied}\}. \tag{16}$$

If a vector X of unknowns is formulated, where

$$X = [f_{1x} f_{1y} \cdots f_{n_w x} f_{n_w y} F_1 \cdots F_{n_w} \ddot{s}\dot{s}^2], \tag{17}$$

then mathematical program (16) has linear objective and constraints, except (13) which is convex though.

2.3. *Dynamic modeling*

The simplying assumptions made to obtain the dynamic model of the mobile robot are:

- No slipping: rolling compatibility conditions are satisfied.
- No skidding: $|v(s)| \le v_{\text{skid}}(s)$.
- The rotational kinetic energy of the rotating wheels is not considered.
- Frictionless motion (w.r.t. the robot's kinematic mechanism).

The kinetic and potential energies of the moving robot are:

$$K = \tfrac{1}{2}\langle \vec{\omega}, I\vec{\omega} \rangle + \tfrac{1}{2}m\dot{s}^2, \tag{18}$$

$$P = mgz(s). \tag{19}$$

where I is the moment of inertia of the robot around G, m is its mass, g is the gravity vector and $z(s)$ is the height of the center of gravity w.r.t. a world coordinate frame.

Since $\vec{\omega}(s) = \dot{s}f(s)\vec{c}$, where $f(s)$ is the curvature at s, the kinetic energy becomes

$$K(s, \dot{s}) = \tfrac{1}{2}I_c(s)f^2(s)\dot{s}^2 + \tfrac{1}{2}m\dot{s}^2, \tag{20}$$

where $I_c(s) = \langle \vec{c}, I\vec{c} \rangle$ is the reflected inertia matrix. The Lagrangian is

$$L(s, \dot{s}) = K - P = \tfrac{1}{2}I_c(s)f^2(s)\dot{s}^2 + \tfrac{1}{2}m\dot{s}^2 - mgz(s).$$

The Euler–Lagrange equations, assuming no nonconservative forces, such as friction, are:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{s}} - \frac{\partial L}{\partial s} = u,$$

where u is the steering force of the robot tangent to $r(s)$ at every instant. Using (20), the dynamic equation is

$$(m + I_c(s)f^2(s(t)))\ddot{s}(t) + (\tfrac{1}{2}I_c'(s)f^2(s) + I_c(s) \times f(s(t))f'(s(t)))\dot{s}^2(t) + mgz(s) = u(t), \tag{21}$$

where $(\cdot)' \triangleq \mathrm{d}(\cdot)/\mathrm{d}s$. An equivalent useful formulation may be obtained using parameter s as an independent variable instead of time t. This is accomplished by introducing $v(s) \triangleq \mathrm{d}s/\mathrm{d}t(s)$. Therefore

$$(m + I_c(s)f^2(s))v(s)v'(s) + (\tfrac{1}{2}I_c'(s)f^2(s) + I_c(s) \times f(s)f'(s))v^2(s) + mgz(s) = u(s). \tag{22}$$

The steering force u is the sum of the projections of the n_w actuator forces on the tangent of $r(s)$

$$u(t) = u(s(t)) = \sum_{i=1}^{n_w} u_i \sin(\phi_i^0(s(t))). \tag{23}$$

The p_y^0 components of u_is are absorbed by friction. Usually, the steering angles ϕ_i^0 of the wheels have a limited range of deviation around 90°. Therefore

$$c_1 \le |\sin(\phi_i^0(s(t)))| \le 1, \tag{24}$$

where c_1 is a positive constant. From (15), (23) and (24)

$$-U_2 \le u \le U_1, \tag{25}$$

where

$$U_1 = c_1 n_w \mathcal{F}_1, \quad U_2 = c_1 n_w \mathcal{F}_2. \tag{26}$$

This is an acceptable approximation of the input space because it is on the safe side.

2.4. *Collision avoidance constraints*

Collision avoidance is guaranteed if the distance $d(s, t)$ (Gilbert and Johnson, 1985) between the robot and the object is greater than a safety positive constant d^0, i.e.

$$d(s,t) = \min_{i,j} \{\|z_i - z_j\| : z_i \in C_r(s), z_j \in C_o(t)\} \ge d^0, \quad \forall s, t, \tag{27}$$

where

$$C_r(s) = \{x/x \in \Re^3 \ni A_r R_r^{-1}(s)x \le b_r - A_r R_r^{-1}(s) T_r(s),\ A_r \in \Re^{m\times 3},\ b_r \in \Re^m\}, \tag{28}$$

$$C_o(t) = \{y/y \in \Re^3 \ni A_o R_o^{-1}(t) y \le b_o A_o R_o^{-1}(t) T_o(t),\ A_o \in \Re^{l\times 3},\ b_o \in \Re^l\}, \tag{29}$$

are convex polyhedra representing the convex hulls of the mobile robot and the moving obstacle, respectively. (A_r, b_r) and (A_o, b_o) are the parameters that define the convex polyhedron description of the robot and the object, respectively, with respect to their fixed coordinate frame. R_r, R_o, T_r, T_o represent the rotation and translation of the frames of the robot and the object with respect to the world frame and must be known in order to compute $d(s, t)$. $R_r(s)$, $T_r(s)$ can be easily derived from $r(s)$. $R_o(t)$, $T_o(t)$ are not well known and are estimated based on sensory input. Distance $d(s, t)$ can be computed by a mathematical program of the form:

$$d(t) = \min_{x,y} \|x - y\|_a \quad \text{s.t } A_r(t)x(t) \le b_r(t), \quad A_o(t)y(t) \le b_o(t). \tag{30}$$

Normally one would choose the Euclidean distance ($a = 2$) to represent the distance, and have to solve a quadratic programming type of problem. However if $a = 1$ or ∞ then a linear programming type of problem can be formulated and gain in terms of computational efficiency.

2.5. *Mathematical problem statement*

The dynamic equations and the constraints are summarized as follows:

System (equation (22))

$$\dot{x}(t) = A(x(t)) + B(x(t))u(s), \tag{31}$$

where $x(t) = [s(t)v(t)]^T$, with $v(t) = \dot{s}(t)$

$$A(x) = \begin{bmatrix} v(t) \\ -\dfrac{\frac{1}{2}I_t'(s)f^2(s) + If(s)f'(s)}{m + I_t(s)f^2(s)} v^2(t) \end{bmatrix} \tag{32}$$

and

$$B(x) = \begin{bmatrix} 0 \\ \dfrac{1}{m + I_t(s)f^2(s)} \end{bmatrix}. \tag{33}$$

Initial–final conditions

$$x(s_0) = [0 \quad v_0]', \quad x(s_f) = [\text{free } v_f]'. \tag{34}$$

Input constraint (equation (25))

$$-U_2 \le u(s) \le U_1, \tag{35}$$

State constraint (equation (16))

$$|v(s)| \le v_{skid}(s), \tag{36}$$

Collision avoidance constraint (equation (27))

$$d_o - d(s, t) \le 0. \tag{37}$$

The final equality constraint (34) can be imposed using inequalities. Consider as upper and lower velocity bounds the two velocity arcs starting from (s_f, v_f) and going backwards with input $-U_2$, U_1 until they meet $(v_{skid}(s)$, $v(s) = 0)$, respectively. Thus $v_{max}(s)$ and $v_{min}(s)$ are constructed. This is demonstrated in Fig. 4. Thus, a new set of inequalities

$$0 \le v_{min}(s) \le |v(s)| \le v_{max}(s), \tag{38}$$

are used to satisfy (34) and (36).

The problem is to find a feedback control of the form $u(s, v, t)$ such that (35), (37), and (38) are satisfied.

3. THE POTENTIAL FIELDS STRATEGY (PFS)

The potential fields approach does not generically address the performance issue. Stability, in the sense of collision avoidance and convergence to the goal state, is the main issue. Furthermore, being a local method, it is computationally efficient but on the other hand it does not handle constraints well. The main problem with a naive application of this approach is the existence of local minima in the overall potential field. A considerable research effort has been done Rimon and Koditschek (1989) on constructing globally converging fields, but this has been achieved only for special categories of shapes of objects (e.g. "star-like"). The above issues become more complicated if

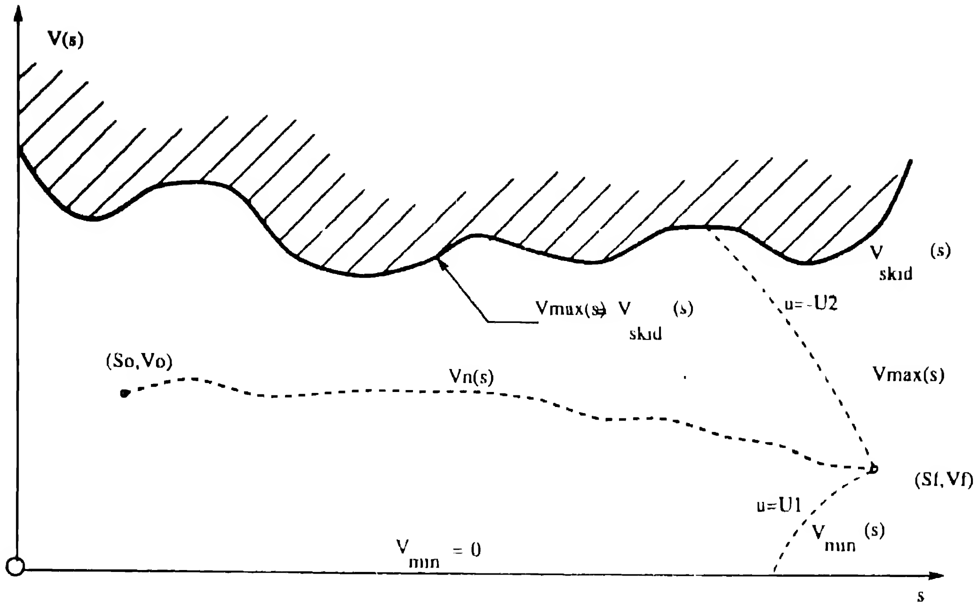


FIG. 4 Feasible velocities space

the objects move and the system becomes non-autonomous.

In the problem at hands, the application of potential fields is done in the (s, t) space since the problem has been reduced to a unidimensional one by considering motions only along the off-line trajectory $r(s)$. A Lyapunov function approach was adopted. In order to aim for both stability and performance, two generic states are considered for the case of having only one moving object:

- Alarming state. Collision is possible unless action is taken. The inverse of the distance between the robot and the moving obstacle is penalized. The attractive goal is state ($s = s_f$, $v = v_f = 0$).
- Non-alarming state. The main task is to track the nominal plan. Therefore the attractive state is $s_n(t)$, $v_n(t)$.

This decomposition demonstrates the necessity of development of a fast but generic scheme that classifies the current state of the system as alarming or not.

3.1. *Fast collision prediction—the alarm function*

In our previous developments the Minimum Interference Strategy (MIS) and the Optimal Control Strategy (OC) Kyriakopoulos and Saridis (1990b, 1991, 1992b), have been proposed. Those are global methods and require the accurate knowledge of parameters such as collision time (t_c), collision point (s_c), etc. In the case of PFS, collision prediction is only required to test if the local configurations and velocities of the mobile robot and the moving obstacle may lead to a collision.

Consider Fig. 5. A mobile robot is moving along its Cartesian path and has an instantaneous translational velocity $\vec{v}_R = \vec{v}_r + v_R^z\vec{k}$, while a moving obstacle is detected to have an instantaneous translational velocity† $\vec{v}_O = \vec{v}_o + v_O^z\vec{k}$. Configuration (a) is considered as alarming because the shaded area (volume) which is the interaction of the swept areas (volumes) of the polyhedra along their instantaneous translational motion lines lie on the side of positive direction of both velocity vectors. Configuration (b) is considered as non-alarming because the shaded area lies on the negative side for velocity vector v_o. A mathematical methodology to construct an alarm function $F_a(s, t)$ indicating the "degree of alarm" of a configuration is presented below.

Consider the convex polyhedral descriptions of the mobile robot and the moving obstacle, respectively:

$$A_r x_r \leq B_r, \tag{39}$$

$$A_o \cdot x_o \leq B_o, \tag{40}$$

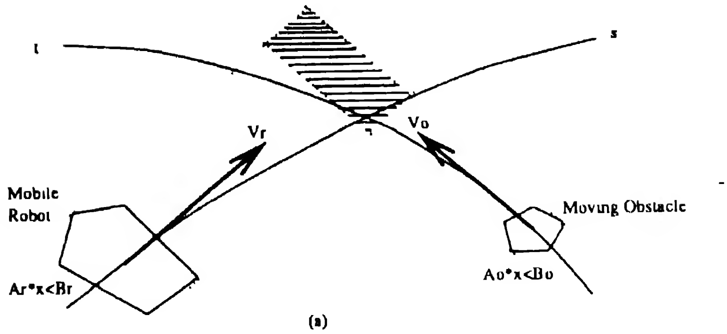
x_r, $x_o \in \Re^2$. Notice that (A_r, B_r) are functions of s, while (A_o, B_o) are functions of t. The swept areas (volumes) of the mobile robot and the moving obstacle under a horizontal translation can be parametrized by

$$x_r' = x_r + \lambda v_r, \tag{41}$$

$$x_o' = x_o + \mu v_o, \tag{42}$$

where x_r, x_o satisfy (39) and (40), respectively. Their intersection (shaded area of Figs 5a and b)

† Obviously $\vec{v}_r$, $\vec{v}_o$ are the horizontal components of the velocities of the mobile robot and the moving obstacle, respectively.



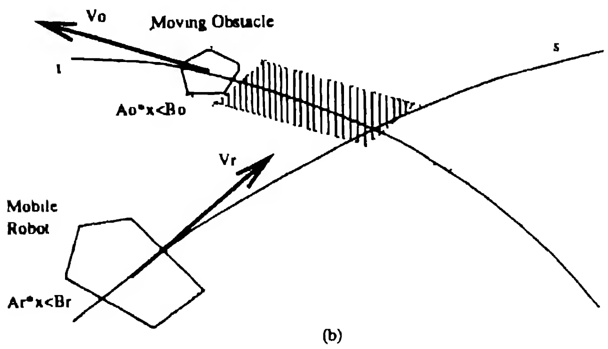


FIG 5 (a) Alarming configuration (b) Non-alarming configuration

is given by

$$x_r + \lambda v_r = x_o + \mu v_o. \tag{43}$$

Introducing (43) to (39) and (40) the description of the shaded area S is obtained

$$S = \{X \in \Re^6 \ni AX \le B\}, \tag{44}$$

where

$$A = \begin{bmatrix} A_r & 0 & A_r v_o & -A_r v_r \\ 0 & A_o & -A_o v_o & A_o v_r \end{bmatrix}, \quad X = \begin{bmatrix} x_o \\ x_r \\ \mu \\ \lambda \end{bmatrix},$$

$$B = \begin{bmatrix} B_r \\ B_o \end{bmatrix}. \tag{45}$$

The configuration is alarming only if;

$$\lambda_{max} \triangleq \max_{X \in S} \lambda > 0, \tag{46}$$

and

$$\mu_{max} \triangleq \max_{X \in S} \mu > 0. \tag{47}$$

Therefore, in order to determine λ_{max}, μ_{max}, two simple linear programs should be solved. Notice, that λ_{max}, μ_{max} are functions of (s, t) because they depend on $A_r(s)$, $B_r(s)$, $A_o(t)$, $B_o(t)$.

In addition to the above, if

$$\lambda_{max} > \lambda_{alarm}, \tag{48}$$

or

$$\mu_{max} > \mu_{alarm}, \tag{49}$$

where λ_{alarm}, μ_{alarm} are large positive constants, then the configuration should not be considered as alarming.

Thus the alarm function is defined as

$$F_a(s, t) \triangleq u_s(\lambda_{max}(s, t))u_s(\lambda_{alarm} - \lambda_{max}(s, t)) \times u_s(\mu_{max}(s, t))u_s(\mu_{alarm} - \mu_{max}(s, t)), \tag{50}$$

where $u_s(x)$ is the step function Notice that

$$F_a(s, t) = \begin{cases} 1 & \text{alarming rate} \\ 0 & \text{non-alarming state} \end{cases} \tag{51}$$

3.2 $F_a = 0$ *Non-alarming state*

The considered dynamic system is

$$\frac{ds}{dt} = v(s), \tag{52}$$

$$\mathcal{M}(s)\frac{dv}{dt} = -a(s)v^2(t) + u(t), \tag{53}$$

where

$$\mathcal{M}(s) = m + I_c(s)f^2(s),$$

$$a(s) = \tfrac{1}{2}I_c'(s)f^2(s) + I_c(s)f(s)f'(s) = \frac{1}{2}\frac{d\mathcal{M}(s)}{ds}.$$

Consider the proposed Lyapunov function candidate

$$\mathcal{V}(t, s, v) = \tfrac{1}{2}\mathcal{M}(s)(v - v_n(t))^2 + \tfrac{1}{2}K_p(s - s_n(t))^2, \tag{54}$$

where $s_n(t)$, $v_n(t)$ is the nominal plan, and $K_p > 0$. The idea behind this function is to try to lead the system as close as possible to the nominal plan since there is not any alarm.

In general, $\mathcal{V}(t, s, v) \Leftrightarrow (s, v) = (s_n, v_n)$. Furthermore, $\mathcal{V}(t, s, v)$ is continuously differentiable and bounded from below by a decrescent function. Therefore it is a valid Lyapunov function candidate. Its time derivative is:

$$\frac{d\mathcal{V}}{dt} = \frac{\partial \mathcal{V}}{\partial t} + \frac{\partial \mathcal{V}}{\partial s} v + \frac{\partial \mathcal{V}}{\partial v} \dot{v}, \tag{55}$$

and if (54) is substituted, then

$$\frac{d\mathcal{V}}{dt} = (v - u_n)[K_p(s - s_n) - a(s)vv_n - \mathcal{M}(s)\dot{v}_n + u]. \tag{56}$$

For

$$u(t, s, v) = -K_p(s - s_n(t)) - K_v(v - v_n(t)) + a(s)vv_n(t) + \mathcal{M}(s)\dot{v}_n, \tag{57}$$

$(K_v > 0)$ equation (56) becomes

$$\frac{d\mathcal{V}}{dt} = -K_v(v - v_n)^2. \tag{58}$$

Using a well-known result by Hale (1980) (see Appendix B) and the fact that $s_n(t)$, $v_n(t)$ are uniformly bounded in t,

$$\lim_{t\to\infty} (v(t) - v_n(t)) = 0.$$

Notice that

$$s_n(t) = s_f, \quad v_n(t) = \dot{v}_n(t) = 0 \quad t > T, \tag{59}$$

where T is the motion time of the nominal plan. Since $u(t, s, v)$ is continuous and uniformly bounded in t then v, $\dot{v}$ are continuous and uniformly bounded, and using a well-known result (see Appendix B)

$$\lim_{t\to\infty} (\dot{v}(t) - \dot{v}_n(t)) = 0.$$

Thus, from (53) and (57) we deduce that

$$\lim_{t\to\infty} (s(t) - s_n(t)) = 0.$$

3.3. $F_a = 1$; *Alarming state*

In this case, both collision avoidance and convergence to the final state $(s_f, v_f = 0)$ should be satisfied. Therefore the proposed Lyapunov function candidate is

$$\mathcal{V}(t, s, v) = \tfrac{1}{2}\mathcal{M}(s)v^2 + \tfrac{1}{2}(K_p + cK_v)\,\Delta s^2 + \tfrac{1}{2}c\mathcal{M}(s)\,\Delta s v + G(d(s, t)), \tag{60}$$

where $\Delta s = (s - s_f)$, K_p, $K_v > 0$. $\mathcal{V}(t, s, v)$ will be everywhere positive if

$$0 < c < \frac{K_v + \sqrt{K_v^2 + M_0 K_p}}{\frac{M}{2}} \leq \frac{K_v + \sqrt{K_v^2 + \mathcal{M}K_p}}{\frac{\mathcal{M}}{2}}, \tag{61}$$

where $M_0 \leq \mathcal{M}(s) \leq M$, and

$$G(d) = \begin{cases} (d - D) + (D + d_0)\ln\dfrac{D + d_0}{d + d_0} & d \leq D \\ 0 & d > D, \end{cases} \tag{62}$$

with

$$G'(d) = \begin{cases} 1 - \dfrac{D + d_0}{d + d_0} & d \leq D \\ 0 & d > D, \end{cases} \tag{63}$$

$G(d) \geq 0$, (Fig. 6) continuously differentiable,

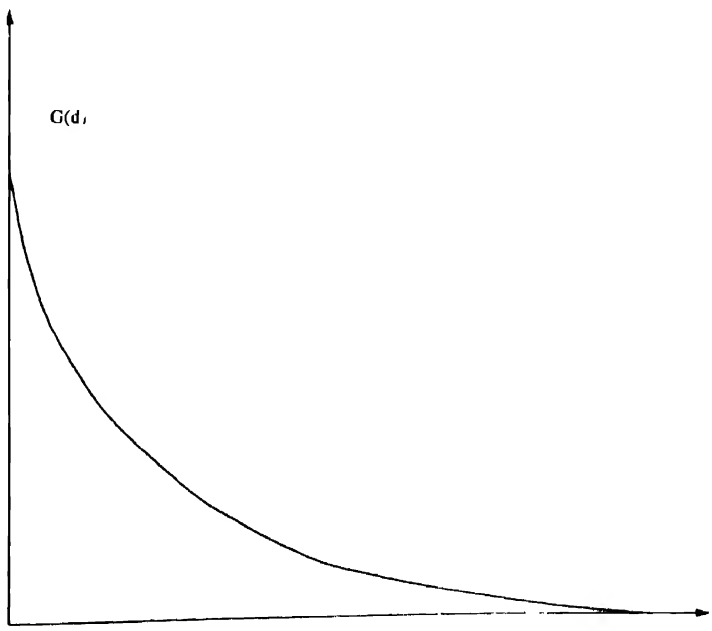


Fig. 6 Structure of $G(d)$.

and decreasing ($G'(d) \le 0$) w.r.t. its argument d. D is the largest distance for which we are concerned, and d_0 is appropriately chosen to define $G(0)$. The idea behind such a construction of $\mathcal{V}(t, s, v)$ is to try to lead the state of the system to this of the desired state, and at the same time to penalize proximity with the moving obstacle when the configuration of the system may lead to a collision. $\mathcal{V}(t, s, v)$ is continuous and bounded from below by a decrescent function. Therefore it is a valid Lyapunov function candidate.

In the following analysis $\partial d/\partial s$, $\partial d/\partial t$ are required. Distance function $d(s, t)$ is well known to be continuous but not continuously differentiable Gilbert and Johnson (1985). A way to bypass this difficulty is to predict the "singularity" points where the derivative is discontinuous based on the estimate of the motion vector of the moving obstacles and interpolate locally with a signoid function. This approach was used in the case study and did not create any numerical problems. Thus, the notation $\partial d/\partial s$, $\partial d/\partial t$ is going to be used rather loosely here.

$$\begin{aligned}\frac{d\mathcal{V}}{dt} &= G'(d)\frac{\partial d}{\partial t} + a(s)v^3 + (K_p + cK_v)\,\Delta s v \\ &\quad + ca(s)\,\Delta s v^2 + \tfrac{1}{2}\mathcal{M}(s)v^2 + G'(d)\frac{\partial d}{\partial s} + \mathcal{M}(s)v\dot{v} \\ &\quad + \tfrac{1}{2}\mathcal{M}(s)\,\Delta s\dot{v} + (v + \tfrac{1}{2}c\,\Delta s)\mathcal{M}(s)\dot{v} = G'(d)\frac{\partial d}{\partial t} \\ &\quad + a(s)v^3 + (K_p + cK_v)\,\Delta s v + ca(s)\,\Delta s v^2 \\ &\quad + \tfrac{1}{2}\mathcal{M}(s)v^2 + G'(d)\frac{\partial d}{\partial s} + \mathcal{M}(s)v\dot{v} \\ &\quad + \tfrac{1}{2}\mathcal{M}(s)\,\Delta s\dot{v} + (v + \tfrac{1}{2}c\,\Delta s)(-av^2 + v). \end{aligned} \tag{64}$$

If a control

$$u(t, s, v) = -a(s)v^2 - K_p\,\Delta s - 2K_v v - G'(d)\frac{\partial d}{\partial s}, \tag{65}$$

is chosen, then

$$\begin{aligned}\frac{d\mathcal{V}}{dt} &= -\left\{\frac{cK_p}{2}\Delta s^2 + \left(2K_v - \frac{c\mathcal{M}}{2}\right)v^2\right\} \\ &\quad - \frac{c}{2}G'(d)\frac{\partial d}{\partial s}\Delta s - a(s)v^3 + G'(d)\frac{\partial d}{\partial t}. \end{aligned} \tag{66}$$

Notice that if

$$c \le \frac{4K_v}{K_p + M} \le \frac{4K_v}{K_p + \mathcal{M}}, \tag{67}$$

then

$$\begin{aligned}\min\left\{\frac{cK_p}{2}, 2K_v - \frac{c\mathcal{M}}{2}\right\} &= \frac{cK_p}{2} > 0, \\ \max\left\{\frac{cK_p}{2}, 2K_v - \frac{c\mathcal{M}}{2}\right\} &= 2K_v - \frac{c\mathcal{M}}{2}. \end{aligned} \tag{68}$$

It can be easily tested that for K_v, $K_p > 0$ constraint (67) is more constraining than (61). Similarly,

$$\begin{aligned} -a_0 &\le a(s) \\ &= \tfrac{1}{2}I_c'(s)f^2(s) + I_c(s)f(s)f'(s) \le a_0 \\ &= \tfrac{1}{2}I'_{c_{\max}}f^2_{\max} + I_{c_{\max}}f_{\max}f'_{\max}, \quad -S_0 \le \frac{\partial d}{\partial s} \le S_0 \\ &= 1 + \delta_r f_{\max}, \end{aligned} \tag{69}$$

where $I_c(s)$ is the reflected moment of inertia of the mobile robot at point s, $I_{c_{\max}} \triangleq \max_s I_c(s)$, $I'_{c_{\max}} \triangleq \max_s I_c'(s)$, $f_{\max} \triangleq \max_s f(s)$, $f'_{\max} \triangleq \max_s f'(s)$ and δ_r the diameter of the robots convex hull.

The required analysis is very difficult because the system is nonautonomous, and the distance function is not explicitly known as a function of (s, t). A local analysis assuming that

$$\begin{aligned} -s_f \le \Delta s &\le \Delta s_t < 0, \\ |v| \le v_{\max} &\triangleq \max v_{\max}(s), \end{aligned} \tag{70}$$

will be performed.

(A) *Stationary obstacle*. $\dfrac{\partial d}{\partial t} = 0$.

In this case

$$\begin{aligned}\frac{d\mathcal{V}}{dt} &= -\left\{\frac{cK_p}{2}\Delta s^2 + \left(2K_v - \frac{c\mathcal{M}}{2}\right)v^2\right\} \\ &\quad - \frac{c}{2}G'(d)\frac{\partial d}{\partial s}\Delta s - a(s)v^3. \end{aligned} \tag{71}$$

From Lasalles' invariance principle, is

$$\begin{aligned}\mathcal{W} \triangleq \Big\{(s, v) : \dot{V}(s, v) &= -\left\{\frac{cK_p}{2}\Delta s^2 + \left(2K_v - \frac{c\mathcal{M}}{2}\right)v^2\right\} \\ &\quad - \frac{c}{2}G'(d)\frac{\partial d}{\partial s}\Delta s - a(s)v^3 = 0\Big\}, \end{aligned} \tag{72}$$

then the state of the system asymptotically goes to the largest invariant set of $\mathcal{W}$, denoted as $\{\mathcal{W}\}$, i.e. (s_e, v_e) satisfying

$$\begin{aligned} v_e &= 0, \\ K_p(s_e - s_f) &= -G(d(s_e))\frac{\partial d}{\partial s} \end{aligned} \tag{73}$$

Obviously, the equilibrium point $(s_e, v_e) \in \{\mathcal{W}\}$ depends on the initial state. For example, if $\partial d(s)/\partial s > 0\ \forall s \in [s_p, s_f]$ and $s_0 \in [s_p, s_f]$ then $(s_e, v_e) = (s_f, 0) \notin \{\mathcal{W}\}$.

(B) *Moving obstacle.* $\dfrac{\partial \mathrm{d}}{\partial \mathrm{t}} \neq 0$.

The analysis of this case is very difficult for the general case and can be only local Khalil and Kokotovic (1991). The following analysis is based on the assumption that the velocity of the moving obstacle is bounded. Thus,

$$-V \leqq \frac{\partial d}{\partial t} \leq V, \tag{74}$$

then

$$G'(d)\frac{\partial d}{\partial t} \leq \cdot\ G'(0)V. \tag{75}$$

The requirement that

$$\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} = -\left\{\frac{cK_p}{2}\Delta s^2 + \left(2K_v - \frac{c\mathcal{M}}{2}\right)v^2\right\}$$
$$-\frac{c}{2}G'(d)\frac{\partial d}{\partial s}\Delta s - a(s)v^3 + G'(d)\frac{\partial d}{\partial t} < 0, \tag{76}$$

is satisfied if

$$\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} \leq \max \dot{V}(s, v, d) = -\frac{cK_p}{2}\Delta s_f^2 - \frac{c}{2}G'(0)S_0 s_f$$
$$+ a_0 v^3_{\max} - G'(0)V < 0, \tag{77}$$

which is true for

$$K_v \geq \frac{a_0 v^3_{\max} - G'(0)V}{2\,\Delta s_f^2}, \tag{78}$$

and

$$K_n \quad \frac{(a_0 v^3_{\max} \quad G'(0)V)M - 4K_v G'(0)S_0 s_f}{2K_v\,\Delta s_f^2 - a_0 v^3_{\max} + G'(0)V} \tag{79}$$

as it is shown in Appendix C.

Depending upon the assumptions for $\partial d/\partial t$, several conclusions about the stability properties can be drawn from inequality (76) based on the development of Khalil and Kokotovic (1991). For example, if in addition to the boundedness of $\partial d/\partial t$ we have $\partial d/\partial t \to 0$ as $t \to 0$ then we can conclude that $(s(t), v(t)) \to (s_e, v_e) \in \{W\}$.

The algorithm for the Potential Fields Strategy is:

Algorithm for PFS.

Step 1. *Collision prediction*

Solve (46) and (47) for $\lambda_{\max}$, $\mu_{\max}$, respectively;

Find $F_a(s, t)$ from (50);

if $F_a(s, t) = 1$ GOTO 2;

else $(F_a(s, t) =$ GOTO 3.

Step 2. *Collison avoidance*

Find minimum distance $d(s, t)$;

Find $u(t, s, v)$ from (65);

GOTO 1.

Step 3. *Tracking stage*

Find $u(t, s, v)$ from (57);

GOTO 1.

3.4. *Implementation issues*

The control input $u(t, s, v)$ obtained from (57) or (65) may violate the limits imposed by (25), $v_{\max}(s)$ and $v_{\min}(s)$. This can be avoided by imposing

$u(t, s, v) :=$

$$\begin{cases} \max\{U_1, u(t, s, v), \mathcal{M}(s)v_{\max}(s) \times v'_{\max}(s) \\ \quad + a(s)v^2_{\max}(s)\}, \quad u(t, s, v) \geq 0, \\ \min\{-U_2, u(t, s, v), \mathcal{M}(s)v_{\min}(s) \times v'_{\min}(s) \\ \quad + a(s)v^2_{\min}(s)\}, \quad u(t, s, v) \leq 0. \end{cases} \tag{80}$$

Obviously, the system may approach the limits of the feasible space but it will never violate it.

If $d(s, t)$ becomes very small then the input u becomes very large. However in reality $-U_2 \leq u \leq U_1$, indicating that collision avoidance is not always guaranteed because of the limited capacity of the actuators.

After those practical constraints are considered, stability analysis becomes a formidable task. However, the stability for the unbounded case shows that the proposed strategy: (a) tries to classify the moving obstacles as disrupting or not, (b) tends to avoid the disrupting ones, and finally (c) when avoidance has been achieved, tracking of the nominal plan is its main purpose. The simulation results of the next section demonstrate those stages.

3.5. *Inverse dynamics*

The control force u given by equation (80) is actually the sum of the projections of the individual control forces u_i, given their linear combination u, is by solving the following nonlinear programming problem

$$\max_{i} \min_{1\ n_w} \eta^2 F_i^2 - (f_{ix}^2 + f_{iy}^2)$$
$$\text{s.t} \qquad \text{equation (11)–(15)} \tag{81}$$
$$\sum_{i=1}^{n_w} u_i \sin(\phi_i^0(s(t))) = u \text{ (given)},$$

that can be reformulated as

$$\max_{x} w$$
$$w \leq \eta^2 F_i^2 - (f_{ix}^2 + f_{iy}^2), \quad i = 1, \ldots\ n_w,$$
$$w \geq 0 \tag{82}$$
$$\text{equation (11)–(15)}$$
$$\sum_{i=1}^{n_w} u_i \times \sin(\phi_i^0(s(t)))$$
$$= u \text{ (given in equation (80))},$$

where $X = [f_{1x} f_{1y} \cdots f_{n_w x} f_{n_w y} \quad F_1 \cdots F_{n_w} \ddot{s} \dot{s}^2 w]$. The idea behind this optimization is to try to allocate the forces u_i in a such a way so that saturation of the available friction at each wheel is avoided.

4 SIMULATION RESULTS

In this section a case study is presented. A mobile robot and a moving obstacle with geometric shapes, moving in the same environment (Fig. 7). The shape of both the mobile robot and the moving obstacle is rectangle with dimensions $0.3\,\text{m} \times 0.52\,\text{m}$ and $0.28\,\text{m} \times 0.28\,\text{m}$, respectively. The scenario is that when the mobile robot is about to start moving, an obstacle with kinematic parameters $x_0 = 4.0\,\text{m}$, $y_0 = 7.0\,\text{m}$, $v_x = 0.04\,\text{m sec}^{-1}$, $v_y = -0.075\,\text{m sec}^{-1}$, $a_x = 0.094\,\text{m sec}^{-2}$, $v_y = -0.041\,\text{m sec}^{-2}$, is going to collide with it under the current plan.

The mobile robot has parameters:

Mass (M): 60 kg.
Inertia (I_{zz}): $32\,\text{kg m}^2$.
Maximum accelerating force (U_1): 140 N.
Minimum decelerating force (U_2): -60 N.
Maximum velocity (V_{max}): $8\,\text{m sec}^{-2}$.
Wheel–floor friction coefficient (η): ~ 0.12.

It has the task of going from configuration A to configuration B within $T = 12.1753$ sec. An offline path planning stage is done and a path $r(s)$ $0 \le s \le s_f$ with total length $s_f \approx 14.60$ m is computed. The parameters of $r(s)$ are indicated on Fig. 7.

Initial and final velocities are zero ($v_A = v_B = 0$).

The resulting velocity profile from the Potential Fields Strategy (PFS) is plotted with solid line on Fig. 8. The dotted (...) line is the velocity $v_n(s)$ of the nominal plan. Line (− − −) is the maximum velocity $v_{max}(s)$ along $r(s)$. The noisy components of the velocity profile results from the fact that numerical differentiation was actually done to calculate the distance derivatives.

The time functions for PFS ($t_{pfs}(s)$) and the nominal plant ($t_{nom}(s)$) are plotted on Fig. 9. The dotted (...) line is the velocity $t_n(s)$ of the nominal plan. The final motion time of the nominal plan is $T = t_n(s_f) = 12.1753$ sec, while $T_{pfs} = 11.5008$ sec.

5 CONCLUSIONS

PFS has the numerical advantages of a local method. From (57) and (65) the fact that computation load is introduced only from the calculation of terms $F_a(s, t)$, $d(s, t)$ becomes obvious. For $F_a(s, t)$, it was shown in (46), (47) and (50) that two linear programs of six variables (eight for three dimensions) and $m_r + m_o$ constraints (where m_r, m_o the number of the sides of the polygons of robot and obstacle, respectively) must be solved. For $d(s, t)$ as

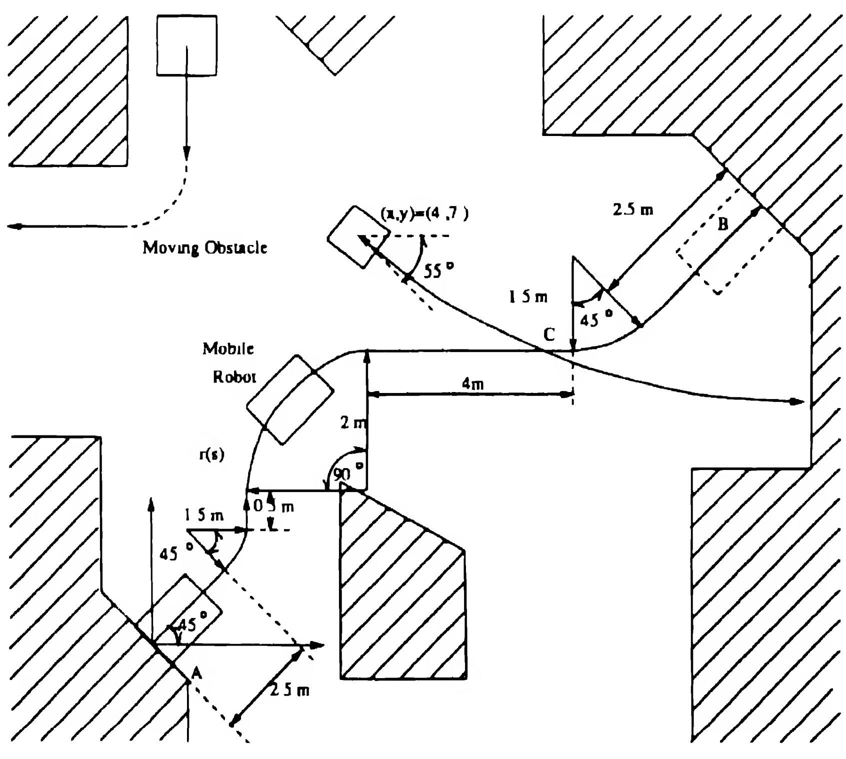


Fig. 7 Environment with a mobile robot and a moving obstacle

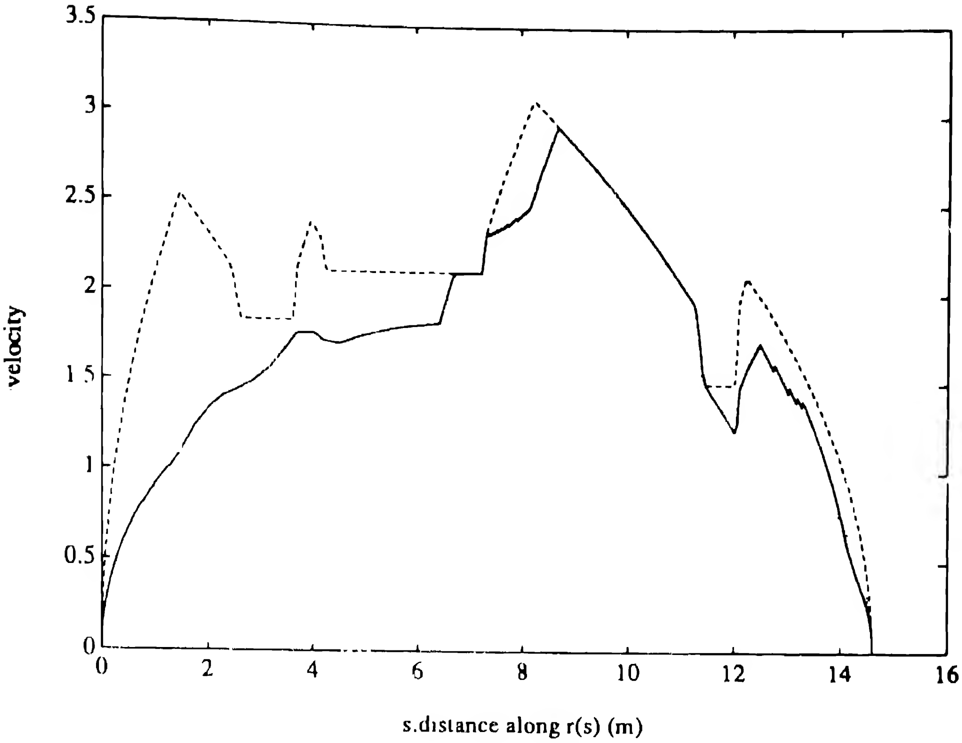


FIG. 8 Velocity profile from PFS (solid)

discussed in Section 2.4, if $\|\cdot\|_1$ or $\|\cdot\|_\infty$ norms are used then a linear program should be solved. Obviously the solution of linear programs of such a small number of variables and constraints does not impose any real-time computational constraint. Therefore this feedback control can be easily recomputed at every sampling interval without any problem.

Currently our efforts are focused at the implementation and comparison of the collision avoidance schemes developed in Kyriakopoulos (1991). It is well understood that the proposal of moving always along the same path will not always prevent collisions. Our research directions include the problem of replanning the shape of $r(s)$ by incorporating the nonholonomic constraints of the rolling motion.

Acknowledgement—This work has been supported by the NASA Center for Intelligent Robotic Systems for Space Exploration (CIRSSE) under the NASA Grant NAGW-1333 The first author has been partially supported by a fellowship of Alexander Onasis Foundation

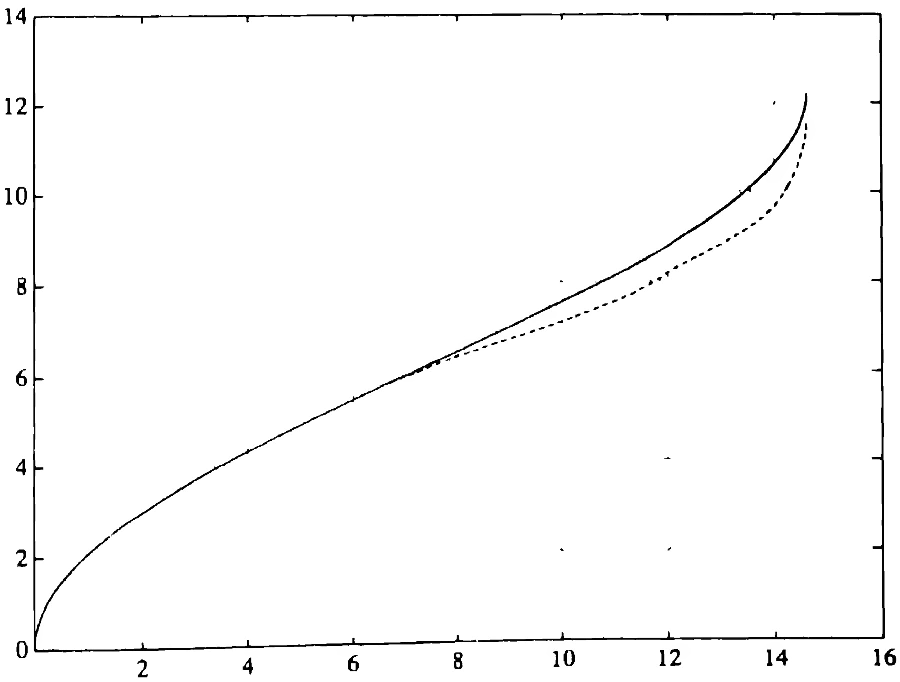


FIG. 9 Time functions for PFS (—) and the nominal plan (solid).

REFERENCES

Alexander, J. and J. Maddocks (1989). On the kinematics of wheeled mobile robots. *Int. J. of Robotics Research*, **8**, 15–27.

Freund, E. and H. Hoyer (1984). Collision avoidance for industrial robots with arbitrary motion. *J. of Robotic Systems*.

Fukimura, K. and H. Samet (1990). Motion planning in a dynamic domain. In *Proc. of the* 1990 *IEEE Int. Conf. on Robotics and Automation*, pp. 324–330.

Gilbert, E. and D. Johnson (1985). Distance functions and their application to robot path planning in the presence of obstacles. *IEEE J. of Robotics and Automation*, pp. 21–30.

Griswold, N. and J. Eem (1990). Control of mobile robots in the presence of moving objects. *IEEE Trans Robotics and Automation*, pp. 263–268.

Hale, J. (1980) *Ordinary Differential Equations* R. E. Krieger, Huntington, NY.

Kant, K. and S. Zucker (1984). Trajectory planning problems, i: Determining velocity along a fixed path. In *Proc. of the IEEE* 8*th Int. Conf. Pattern Recogn.*, pp. 196–198.

Kant, K. and S. Zucker (1986). Toward efficient trajectory planning; The path-velocity decomposition. *Int. J of Robotics Research*, pp. 72–89.

Kant, K. and S. Zucker (1988). Planning collision-free trajectories in time-varying environments: A two-level hierarchy. In *Proc. of the* 1988 *IEEE Int. Conf. on Robotics and Automation*, pp 1644–1649

Khalil, H. K. and P. V. Kokotovic (1991) On stability properties of nonlinear systems with slowly varying inputs *IEEE Trans. Aut. Control*, p. 229.

Khatib, O. (1985). Real time obstacle avoidance for manipulators and mobile robots. In *Proc. of the IEEE Int Conf. on Robotics and Automation*

Kyriakopoulos, K. (1991). A supervisory control strategy for navigation of mobile robots in dynamic environments Ph.D thesis. Technical report, ECSE-RPI.

Kyrikopoulos K. and G. Saridis (1990a). Minimum distance estimation and collision prediction under uncertainty for on-line robotic motion planning. In *Proc of the* 1990 *IFAC World Congress*.

Kyriakopoulos, K. and G. Saridis (1990b) On-line motion planning for mobile robots in non-stationary environments. In *Proc. of* 1990 *IEEE Int. Symposium on Intelligent Control*.

Kyriakopoulos, K. and G. Saridis (1991). Collision avoidance of mobile robots in non-stationary environments. In *Proc. of the* 1991 *Int Conf on Robotics and Automation*.

Kyriakopoulos, K. and G. Saridis (1992a). Minimum distance estimation and collision prediction under uncertainty for on-line robotic motion planning. *Automatica*, **28**, 389–394.

Kyriakopoulos, K. and G. Saridis (1992b). Optimal motion planning for collision avoidance of mobile robots in non-stationary environments. In 1992 *American Control Conf. and Tran. Aut. Control*.

Liu, Y., S. Kuroda, T. Naniwa, H. Noborio and S. Arimoto (1989). A practical algorithm for planning collision free coordinated motion of multiple mobile robots In *Proc. of the* 1989 *IEEE Int. Conf. on Robotics and Automation*, pp. 1427–1432.

Rimon, E. and D. Koditschek (1989). The construction of analytic diffeomorphisms for exact robot navigation on star worlds In *Proc. of the IEEE Int Conf on Robotics and Automation*, pp. 21–26.

Reif, J. and M. Sharir (1985). Motion planning in the presence of moving obstacles. Technical Report TR-06-85, Harvard University, Center for Research in Computing Technology.

Shih, C, T. Lee and W. Gruver (1990). Motion planning with time-varying polyhedral obstacles based on graph search and mathematical programming. In *Proc. of the* 1990 *IEEE Int. Conf. on Robotics and Automation*, pp. 331–337.

Tournassoud, P. (1986). A strategy for obstacle avoidance and its applications to multi-robot systems. In *Proc. of the* 1988 *IEEE Int. Conf. on Robotics and Automation*, pp. 1224–1229.

Wu, C. and C. Jou (1988). Design of a controlled spatial curve trajectory for robot manipulators. In *Proc. of the* 27*th Conf. on Decision and Control*, pp. 161–166.

APPENDIX A. ROBOT INVERSE DYNAMICS

If it is known that the object has, at instant t, translational velocity $v(t)$, orientation angle Θ and rotational velocity $\omega(t) = \dot{\Theta}(t)$ (Fig. 2). Then the instantaneous angular velocity and steering angle of the wheel i, are (Alexander and Maddocks (1989)):

$$\dot{\theta}_i = \frac{1}{r} \times \sqrt{|v(t)|^2 + |x_i^B|^2 (\omega(t))^2 - 2 \times |v(t)|\,|x_i^B|\,\omega(t)\cos(\beta - \alpha_i - \Theta(s(t)))}, \tag{A.1}$$

$$\phi_i^B = \tan^{-1}\left(\frac{|v(t)|\sin\beta - |x_i^B|\,\omega(t)\sin(\alpha_i + \Theta(s(t)))}{|v(t)|\cos\beta - |x_i^B|\,\omega(t)\cos(\alpha_i + \Theta(s(t)))}\right) - \Theta(s(t)). \tag{A.2}$$

There are the inverse kinematics equations for a general mobile robot, i.e. given $v(t)$, $\Theta(t)$, $\dot{\Theta}(t)$ equation (A.2) gives the angular velocities $\dot{\theta}_i(t)$ and steering angles $\phi_i^B(t)$ of every wheel i

APPENDIX B RESULTS NEEDED IN SECTION 3.2

Result 1 Consider the system

$$\dot{x} = f(t, x) \tag{B.1}$$

Let $\mathcal{V}$ be a Lyapunov function candidate for (B.5) Suppose that

$$\dot{\mathcal{V}}(t, x) \le -\mathcal{W}(x) \le 0 \quad \forall t \ge t_0,\ x \in \Re^n, \tag{B.2}$$

where $\mathcal{W}(x)$ is a continuous function of x Define

$$\mathcal{S} \triangleq \{x / \mathcal{W}(x) = 0\}, \tag{B.3}$$

if

$\forall p \in \Re^n\ \exists$ neighborhood $\mathcal{N}(p) \ni \forall x \in \mathcal{N}(p)$, $f(t, x)$ is uniformly bounded in t,

then

$$x(t) \to \mathcal{S} \text{ as } t \to \infty$$

Result 2 If $f(t)$ *and* $\dot{f}(t)$ are continuous and uniformly bounded then

$$\lim f(t) = 0 \Rightarrow \lim \dot{f}(t) = 0$$

APPENDIX C. INDICATION OF THE VALIDITY OF EQUATIONS (78) AND (79)

From (77) and (67) it suffices that

$$-\frac{cK_p}{2}\Delta s_i^2 - \frac{4K_v}{K_p + M} G'(0) S_0 s_i + a_0 v_{max}^3 - G'(0)V < 0. \tag{C.1}$$

Solving for c and introducing (67)

$$\frac{2}{K_p \Delta s_i^2}\left(a_0 v_{max}^3 - G'(0)V - \frac{4K_v}{K_p + M} G'(0) S_0 s_i\right) \le c \le \frac{4K_v}{K_p + M}. \tag{C.2}$$

This inequality is satisfied if (78) and (79) are satisfied.

Automatica, Vol. 29, No. 2, pp. 323–335, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Adaptive Control of Systems with Backlash*

GANG TAO† and PETAR V. KOKOTOVIĆ†

For systems with unknown backlash an adaptive inverse controller promises to significantly improve system performance.

Key Words—Adaptive backlash inverse; adaptive control; backlash; control system design; nonlinear systems, parameter estimation; stability.

Abstract—We develop a mathematical model of backlash inverse and give a parametrization of the error caused by its estimate. We then design an adaptive backlash inverse controller for unknown plants with backlash and prove the global boundedness of the closed-loop signals. Simulations show major improvements of system performance.

1. INTRODUCTION

SYSTEM COMPONENTS, such as hydraulic servovalves or gears often have dead-zone, hysteresis or backlash characteristics (Krasnosel'skii and Pokrovskii, 1983; Mayergoyz, 1991; Netushil, 1973; Truxal, 1958). These nondifferentiable nonlinearities severely limit overall system performance. Their parameters are often unknown, and an appropriate control strategy for such systems is adaptive. However most adaptive control results are for linear plants or plants with differentiable nonlinearities and are not applicable to nondifferentiable nonlinearities. This situation motivated a research program initiated by two recent papers which deal with dead-zone at the input of a linear part of the plant (Recker *et al.*, 1991; Tao and Kokotović, 1991). In this paper we continue this research direction and solve a more difficult problem with backlash. We propose an adaptive backlash inverse design which guarantees signal boundedness and results in major improvements of system performance.

We consider a plant with a linear part $G(D)$ and a backlash nonlinearity $B(\cdot)$ at its input:

$$y(t) = G(D)[u](t), \quad u(t) = B(v(t)), \quad (1.1)$$

where D is used to denote, as the case may be, the Laplace transform variable or the differential operator $D[x](t) = \dot{x}(t)$ in continuous time, and the z-transform variable or the advance operator $D[x](k) = x(k+1)$ in discrete time. Both $G(D)$ and $B(\cdot)$ are unknown. Only the plant output $y(t)$ is measured and the accessible control input is $v(t)$. The backlash output $u(t)$ is not accessible.

We first develop a backlash inverse for the continuous-time case and derive its discrete-time counterpart. We then design an adaptive control scheme for the discrete-time version of the unknown plant (1.1). In Section 2 we introduce the concept of a backlash inverse $BI(\cdot)$ and investigate its implementation. If the backlash characteristic $B(\cdot)$ were known, then an exact backlash inverse $BI(\cdot)$ inserted between a linear controller and the plant would cancel the effect of backlash. In most applications, the backlash nonlinearity cannot be accurately modeled and only an estimate of its inverse can be implemented. In Section 3, we derive a linear-like parametrization convenient for adaptive control. In Section 4, we use a first-order example to introduce the adaptive backlash inverse design and illustrate performance improvement achieved. Our main result is given in Section 5 where we design the adaptive backlash inverse system for unknown plants of higher relative degree.

2. BACKLASH AND ITS RIGHT INVERSE

In this section we develop a backlash inverse $BI(\cdot)$ and present its implementation to be cascaded with the backlash $B(\cdot)$ at the input of the plant.

2.1. *Backlash characteristic*

The backlash characteristic $B(\cdot)$ with input $v(t)$ and output $u(t)$ is described by two straight lines, upward and downward sides of $B(\cdot)$, connected with horizontal line segments. The upward side of $B(\cdot)$ is active when both $v(t)$ and

* Received 2 January 1992; received in final form 9 June 1992. The original version of this paper was presented at the 4th IFAC Symposium on Adaptive Systems in Control and Signal Processing which was held in Grenoble, France during July 1992. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor C. C. Hang under the direction of Editor P. C. Parks.

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, U.S.A.

$u(t)$ grow:

$$u(t) = m_r(v(t) - c_r), \quad m_r > 0, \quad \text{and} \quad \dot{v}(t) > 0. \tag{2.1}$$

If $\dot{v}(t)$ changes sign at $t = t_1$ and $\dot{v}(t) < 0$ for $t > t_1$, then the upward side becomes inactive and the motion for $t > t_1$ proceeds along the inner segment where

$$u(t) = u(t_1) = \text{constant}. \tag{2.2}$$

The downward side of B(·) is

$$u(t) = m_l(v(t) - c_l), \quad m_l > 0, \quad c_l < c_r, \quad \text{and} \quad \dot{v}(t) < 0. \tag{2.3}$$

It is reached at $t = t_2 > t_1$ when

$$u(t_1) = m_l(v(t_2) - c_l). \tag{2.4}$$

During the motion on the inner segment with $u(t) = u(t_1)$, the following condition is satisfied:

$$\frac{u(t_1)}{m_l} + c_l < v(t) < \frac{u(t_1)}{m_r} + c_r. \tag{2.5}$$

Starting with $t = t_2$, the downward side continues to be active as long as $\dot{v}(t) < 0$. If at $t = t_3 > t_2$ $\dot{v}(t)$ changes its sign again, and $\dot{v}(t) > 0$ for $t > t_3$, then the downward side becomes inactive and the motion proceeds on the inner segment where

$$u(t) = u(t_3) = \text{constant}, \tag{2.6}$$

until at $t = t_4 > t_3$ the upward side becomes active, where t_4 satisfies

$$u(t_3) = m_r(v(t_4) - c_r). \tag{2.7}$$

This narrative description of the backlash characteristic is mathematically modeled by:

$$\dot{u}(t) = \begin{cases} m_r\dot{v}(t) & \text{if } \dot{v}(t) > 0 \text{ and } u(t) = m_r(v(t) - c_r) \\ m_l\dot{v}(t) & \text{if } \dot{v}(t) < 0 \text{ and } u(t) = m_l(v(t) - c_l) \\ & \text{if } m_r(v(t) - c_r) < u(t) < m_l(v(t) - c_l), \\ & \text{of } \dot{v}(t) > 0 \text{ and } u(t) = m_l(v(t) - c_l), \\ & \text{or } \dot{v}(t) < 0 \text{ and } u(t) = m_r(v(t) - c_r), \\ & \text{or } \dot{v}(t) = 0. \end{cases} \tag{2.8}$$

Although this model allows the upward and downward slopes to be different $m_r \neq m_l$ provided that the intersection of the two lines is not in the region of practical interest, we only consider the usual backlash characteristic with two parallel sides $m_r = m_l = m$.

A typical motion on such a characteristic initiated at $t = 0$ with $v(0) = 0$ and $u(0) = 0$ is shown in Fig. 1, where $v(t)$ and $u(t)$ are plotted along two synchronized orthogonal t-axes.

It is useful to treat the model (2.8) as a first-order dynamical system and consider $u(t)$ as its state. With an initial condition $u(0)$, the knowledge of $v(t)$ and $\dot{v}(t)$ uniquely defines $u(t)$ for $t \geq 0$. We will restrict $v(t)$ to be piecewise continuous. Except at points of discontinuity of $v(t)$, its derivative $\dot{v}(t)$ will also be piecewise continuous. We note that $u(t)$ is "more discontinuous" than $v(t)$. For example, even if $v(t)$ is twice differentiable, $\dot{u}(t)$ may still be only piecewise continuous, due to the inner segments on which $\dot{u}(t) = 0$ even though $\dot{v}(t) \neq 0$.

Another important observation concerns the discontinuity of $u(t)$ caused by an inconsistent initialization when the pair $v(0)$, $u(0)$ is not a point on the backlash characteristic in the (v, u)-plane. After a jump in $u(t)$ the pair $v(0^+)$, $u(0^+)$ will uniquely define a point on the backlash characteristic and $v(t)$, $u(t)$ will remain on it thereafter.

Because of the dependence of $\dot{u}(t)$ not only on $v(t)$ and $u(t)$, but also on $\dot{v}(t)$, we think of the input–output mapping from $v(t)$ to $u(t)$ defined by (2.8) as a description of a "relative degree zero" system having a causal right inverse. Next we develop such an inverse and use it as a part of our new controller structure.

2.2. *Backlash inverse*

The most damaging effect of backlash on system performance is the delay corresponding to the time needed to traverse an inner segment of B(·). The ideal backlash inverse BI(·) will make the traverse of this segment instantaneous and thus cancel this undesirable backlash effect. Another undesirable effect of backlash is the information loss occurring on an inner segment when the output $u(t)$ remains constant while the input $v(t)$ continues to change (see the "chopped" $u(t)$ in Fig. 1). These two undesirable effects are eliminated with the backlash inverse BI(·) defined by the following mapping from $u_d(t)$ to $v(t)$:

$$\dot{v}(t) = \begin{cases} \frac{1}{m_r}\dot{u}_d(t) & \text{if } \dot{u}_d(t) > 0 \text{ and } v(t) = \frac{u_d(t)}{m_r} + c_r \\ \frac{1}{m_l}\dot{u}_d(t) & \text{if } \dot{u}_d(t) < 0 \text{ and } v(t) = \frac{u_d(t)}{m_l} + c_l \\ 0 & \text{if } \dot{u}_d(t) = 0 \\ g(t, t) & \text{if } \dot{u}_d(t) > 0 \text{ and } v(t) = \frac{u_d(t)}{m_l} + c_l \\ -g(t, t) & \text{if } \dot{u}_d(t) < 0 \text{ and } v(t) = \frac{u_d(t)}{m_r} + c_r. \end{cases} \tag{2.9}$$

In this definition, the inverse of a horizontal

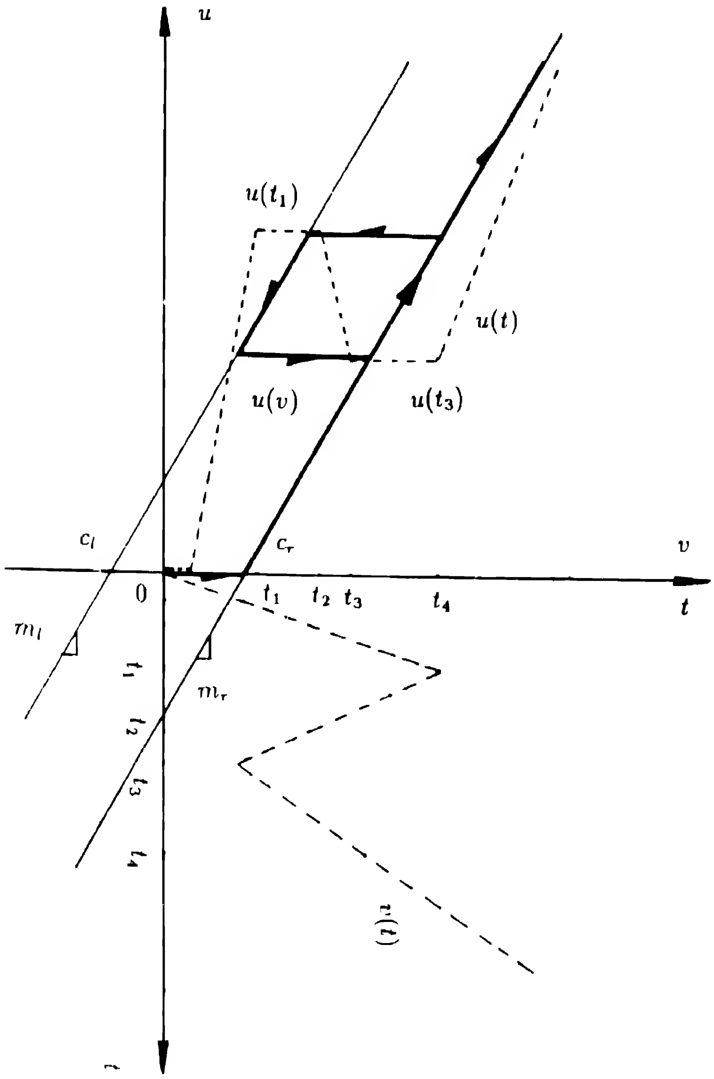


FIG. 1 Backlash model

segment of the backlash characteristic is a vertical jump defined as the time integral of the impulse:

$$g(\tau, t) = \delta(\tau - t)\left(\left(\frac{1}{m_r} - \frac{1}{m_l}\right)u_d(\tau) + c_r - c_l\right), \tag{2.10}$$

where $\delta(t)$ is the Dirac δ-function. Thus an upward jump in the backlash inverse is

$$v(t^+) = v(t^-) + \int_t^{t^+} g(\tau, t)\,\mathrm{d}\tau = v(t^-) + \left(\frac{1}{m_r} - \frac{1}{m_l}\right)u_d(t^-) + c_r - c_l = \frac{u_d(t^-)}{m_r} + c_r. \tag{2.11}$$

The effect of this jump in $\mathrm{BI}(\cdot)$ will be to eliminate the delay caused by a segment in $\mathrm{B}(\cdot)$. In a similar manner the use of (2.9) restores the information that would have been lost in (2.8)

We show this by proving that $\mathrm{BI}(\cdot)$ defined in (2.9) is the right inverse of $\mathrm{B}(\cdot)$ defined by (2.8).

Lemma 2.1 (Backlash inverse). The characteristic $\mathrm{BI}(\cdot)$ defined by (2.9) is the right inverse of the characteristic $\mathrm{B}(\cdot)$ defined by (2.8) in the sense that

$$u_d(t_0) = \mathrm{B}(\mathrm{BI}(u_d(t_0))) \Rightarrow \mathrm{B}(\mathrm{BI}(u_d(t))) = u_d(t), \quad \forall t \geq t_0, \tag{2.12}$$

for any piecewise continuous $u_d(t)$ and any $t_0 \geq 0$.

Proof. Suppose that $\dot{u}_d(t) > 0$ for $t \in [t_0, t_1]$ and some $t_1 > t_0$. First, if $v(t_0) = u_d(t_0)/m_r + c_r$ and $u(t_0) = m_r(v(t_0) - c_r)$, then it follows from (2.9), (2.8) that $\dot{u}(t) = m_r\dot{v}(t) = m_r(\dot{u}_d(t)/m_r) = \dot{u}_d(t)$ for $t \in [t_0, t_1]$ with $u(t_0) = u_d(t_0)$. Hence $\mathrm{B}(\mathrm{BI}(u_d(t))) = u_d(t)$ for any $t \in [t_0, t_1]$. Second, if $v(t_0) = u_d(t_0)/m_l + c_l$ and $u(t_0) = m_l(v(t_0) - c_l)$, then, according to (2.9), $v(t)$ will have a jump at $t = t_0$ so that $v(t) = u_d(t)/m_r + c_r$ for $t = t_0^+$. The

jump in $v(t)$ makes $u(t)$ traverse an inner segment so that $u(t_0^+) = m_r(v(t_0^+) - c_r)$, which reduces to the first case above.

When $\dot{u}_d(t) < 0$ for $t \in [t_0, t_1]$, a similar analysis shows that $B(BI(u_d(t))) = u_d(t)$ for any $t \in [t_0, t_1]$. If $\dot{u}_d(t) = 0$ for $t \in [t_0, t_1]$, then $B(BI(u_d(t))) = u_d(t)$ holds for any $t \in [t_0, t_1]$.

If $\dot{u}_d(t)$ changes the sign at $t = t_1$, then we can repeat the procedure, and show that $B(BI(u_d(t))) = u_d(t)$ for any $t \geq t_0$. ∇

The mapping (2.9)–(2.11) may not define a backlash inverse only if the signal $u_d(t)$ is such that $v(t)$ and $u(t)$ never leave an inner segment. This situation can happen only if $v(0)$, $u(0)$ are initially on an inner segment and $\dot{u}_d(t) = 0$ for $t \geq 0$ or if $\dot{u}_d(t)$ does not change sign but the total increment of $u_d(t)/m_r$ (or decrement of $u_d(t)/m_l$) is insufficient for $v(t)$, $u(t)$ to leave the segment.

As $u_d(t)$ is the design signal at our disposal, the above situation can be remedied. If $u_d(t)$ does not reach t_0 defined in (2.12), then $u_d(t)$, $v(t)$ are initialized by

$$v(t_0^+) = \begin{cases} \dfrac{u_d(t_0)}{m_r} - c_r & \text{if } v(t_0) = \dfrac{u_d(t_0)}{m_l} + c_l \\[2ex] \dfrac{u_d(t_0)}{m_l} + c_l & \text{if } v(t_0) = \dfrac{u_d(t_0)}{m_r} + c_r. \end{cases} \tag{2.13}$$

This will always result in $u_d(t_0^+) = B(BI(u_d(t_0^+)))$ and (2.12) holds thereafter.

When the exact backlash parameters m_r, m_l, c_r, c_l are unknown, we will use the estimated backlash parameters to design an adaptive backlash inverse.

Let $\widehat{m_l}(t)$, $\widehat{c_l}(t)$, $\widehat{m_r}(t)$, $\widehat{c_r}(t)$ be estimates of m_l, m_r, c_l, c_r, and denote the adaptive backlash inverse $\widehat{BI}(\cdot)$ as $BI(\widehat{m_r}(t), \widehat{c_r}(t), \widehat{m_l}(t), \widehat{c_l}(t); \cdot)$.

Graphically, the backlash inverse (2.9)–(2.11) is depicted in Fig. 2 by two straight lines and instantaneous vertical transitions between the lines, where the downward side is

$$v(t) = \frac{u_d(t)}{\widehat{m_l}(t)} + \widehat{c_l}(t), \quad \text{and} \quad \dot{u}_d(t) < 0, \tag{2.14}$$

and the upward side is

$$v(t) = \frac{u_d(t)}{\widehat{m_r}(t)} + \widehat{c_r}(t), \quad \text{and} \quad \dot{u}_d(t) > 0. \tag{2.15}$$

Instantaneous vertical transitions take place whenever $\dot{u}_d(t)$ changes its sign. On the lines $\dot{v}(t) = 0$ whenever $\dot{u}_d(t) = 0$.

In Fig. 2, the motion of $v(t)$, $u_d(t)$ starts with $u_d(0) = v(0) = 0$. At $t = 0$, $v(t)$ vertically moves to the upward side. For $t \in (0, t_1]$, $\dot{u}_d(t) > 0$ does not change sign so the motion stays on the

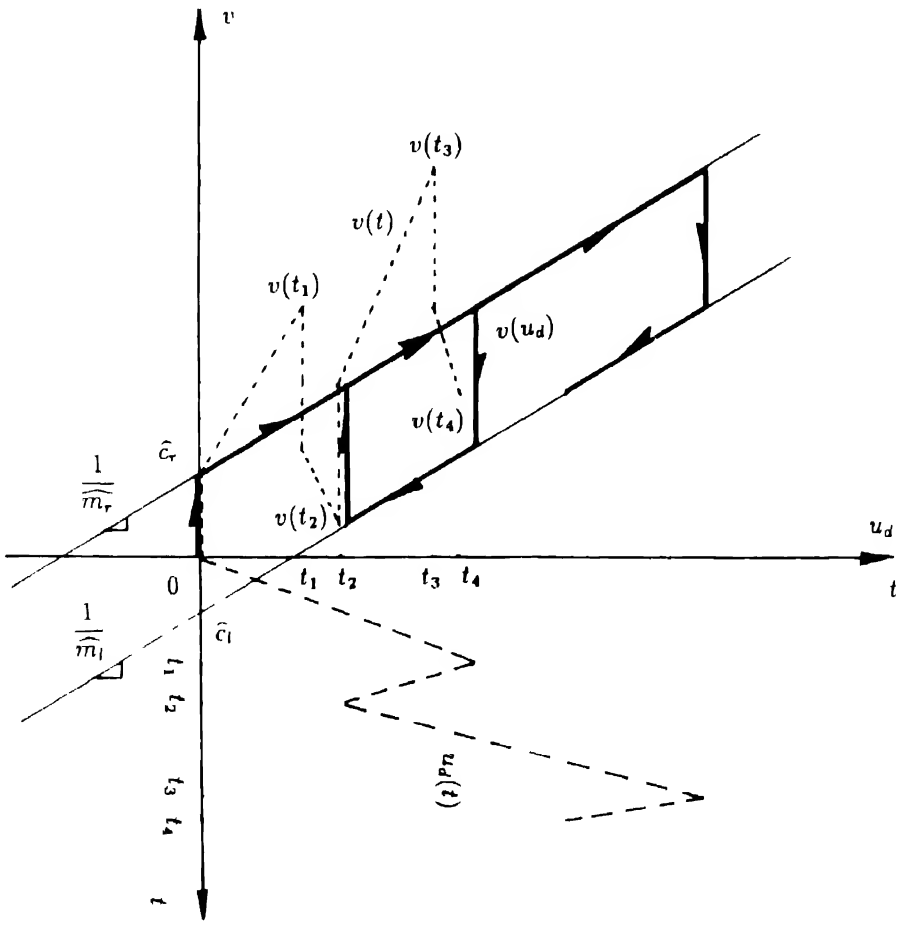


FIG. 2. Backlash inverse.

upward side (2.15). For $t \in (t_1, t_2)$, $\dot{u}_d(t) < 0$, the downward side (2.14) is active. At $t = t_1$, the sign of $\dot{u}_d(t)$ changes causing an instantaneous vertical transition from the upward side to the downward side. Because of the subsequent sign changes of $\dot{u}_d(t)$, two more instantaneous vertical transitions take place; one upward at $t = t_2$ and another one downward at $t = t_3$. The magnitude of the vertical translation of $v(t)$ is equal to the length of the estimated inner segment of $B(\cdot)$. If an exact backlash inverse is used, that is, if $\widehat{m}_l = m_l$, $\widehat{c}_l = c_l$, $\widehat{m}_r = m_r$, and $\widehat{c}_r = c_r$, then, after initialization of the backlash inverse, the backlash output $u(t)$ is equal to $u_d(t)$, that is, $u(t) = B(BI(u_d(t))) = u_d(t)$.

2.3. *Discrete-time representation*

In most applications the accessible control $v(t)$ is piecewise constant, i.e. $v(t) = v(t_k) \triangleq v(k)$ for $t \in [t_k, t_{k+1})$, $k = 0, 1, 2, \ldots$. For such discrete-time applications the backlash model (2.8) is not appropriate because of the discontinuity of signals. However, from the same physical description of backlash given by (2.1)–(2.7), we can obtain the following discrete-time backlash model $u(k) = B(v(k))$:

$$u(k) = \begin{cases} m_l(v(k) - c_l) & \text{for } v(k) \le v_l \\ m_r(v(k) - c_r) & \text{for } v(k) \ge v_r \\ u(k-1) & \text{for } v_l < v(k) < v_r, \end{cases} \tag{2.16}$$

where

$$v_l = \frac{u(k-1)}{m_l} + c_l, \quad v_r = \frac{u(k-1)}{m_r} + c_r. \tag{2.17}$$

Similarly, from (2.14), (2.15) with true backlash parameters, we obtain the exact discrete-time backlash inverse model $v(k) = BI(u_d(k))$ as

$$v(k) = \begin{cases} v(k-1) & \text{for } u_d(k) = u_d(k-1) \\ \dfrac{u_d(k)}{m_l} + c_l & \text{for } u_d(k) < u_d(k-1) \\ \dfrac{u_d(k)}{m_r} + c_r & \text{for } u_d(k) > u_d(k-1). \end{cases} \tag{2.18}$$

The discrete-time version of Lemma 2.1 states that the characteristic $BI(\cdot)$ defined by (2.18) is the right inverse of the characteristic $B(\cdot)$ defined by (2.16) such that

$$u_d(k_0) = B(BI(u_d(k_0))) \Rightarrow B(BI(u_d(k))) = u_d(k), \quad \forall k \ge k_0. \tag{2.19}$$

Similar to the continuous-time case, k_0 is reached when $u_d(k) - u_d(k-1)$ changes sign at k_0. An initialization of $u_d(k)$, $v(k)$ defined by

$$v(k_0^+) = \begin{cases} \dfrac{u_d(k_0)}{m_r} + c_r & \text{if } v(k_0) = \dfrac{u_d(k_0)}{m_l} + c_l \\ \dfrac{u_d(k_0)}{m_l} + c_l & \text{if } v(k_0) = \dfrac{u_d(k_0)}{m_r} + c_r, \end{cases} \tag{2.20}$$

results in $u_d(k_0^+) = B(BI(u_d(k_0^+)))$, where the jump from $v(k_0)$ to $v(k_0^+)$ is instantaneous.

Based on the structure of the backlash inverse (2.18), the discrete-time adaptive backlash inverse $v(k) = \widehat{BI}(u_d(k))$ is

$$v(k) = \begin{cases} v(k-1) & \text{for } u_d(k) = u_d(k-1) \\ \dfrac{u_d(k)}{\widehat{m}_l(k)} + \widehat{c}_l(k) & \text{for } u_d(k) < u_d(k-1) \\ \dfrac{u_d(k)}{\widehat{m}_r(k)} + \widehat{c}_r(k) & \text{for } u_d(k) > u_d(k-1), \end{cases} \tag{2.21}$$

where $\widehat{m}_l(k)$, $\widehat{c}_l(k)$, $\widehat{m}_r(k)$, $\widehat{c}_r(k)$ are the estimates of m_l, c_l, m_r, c_r at $t = t_k$.

3. PARAMETRIZATION

To develop an adaptive law for updating the estimates $\widehat{m}_l(t)$, $\widehat{c}_l(t)$, $\widehat{m}_r(t)$, $\widehat{c}_r(t)$ of the backlash inverse parameters, it is crucial to express the backlash inverse error $u(t) - u_d(t)$ in terms of a parametrizable part and a unparametrizable but bounded part.

To give a compact description for the adaptive backlash inverse, we introduce two indicator functions:

$$\widehat{\chi}_r(t) = \begin{cases} 1 & \text{for } u_d(t), v(t) \text{ on the upward side of } \widehat{BI}(\cdot) \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

$$\widehat{\chi}_l(t) = \begin{cases} 1 & \text{for } u_d(t), v(t) \text{ on the downward side of } \widehat{BI}(\cdot) \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

From (3.1), (3.2), we have that

$$\widehat{\chi}_r(t) + \widehat{\chi}_l(t) = 1, \tag{3.3}$$

$$\widehat{\chi}_l^2(t) = \widehat{\chi}_l(t), \quad \widehat{\chi}_r^2(t) = \widehat{\chi}_r(t), \quad \widehat{\chi}_l(t)\widehat{\chi}_r(t) = 0. \tag{3.4}$$

Using (2.14), (2.15), (3.1)–(3.4), we express $v(t)$

$$v(t) = (\widehat{\chi}_r(t) + \widehat{\chi}_l(t))v(t) = \frac{\widehat{\chi}_r(t)}{\widehat{m}_r(t)} \times (u_d(t) + \widehat{m}_r(t)\widehat{c}_r(t)) + \frac{\widehat{\chi}_l(t)}{\widehat{m}_l(t)} \times (u_d(t) + \widehat{m}_l(t)\widehat{c}_l(t)). \tag{3.5}$$

Similarly, for the backlash $B(\cdot)$, we introduce

three indicator functions $\chi_r(t)$, $\chi_l(t)$ and $\chi_s(t)$, such that

$$\chi_l(t) = \begin{cases} 1 & \text{for } v(t),\ u(t) \text{ on the upward side of } B(\cdot) \\ 0 & \text{otherwise}, \end{cases} \quad (3.6)$$

$$\chi_l(t) = \begin{cases} 1 & \text{for } v(t),\ v(t) \text{ on the downward side of } B(\cdot) \\ 0 & \text{otherwise}, \end{cases} \quad (3.7)$$

$$\chi_s(t) = \begin{cases} 1 & \text{for } v(t),\ u(t) \text{ on an inner segment of } B(\cdot) \\ 0 & \text{otherwise}. \end{cases} \quad (3.8)$$

Using the following obvious relationships:

$$\chi_r(t) + \chi_l(t) + \chi_s(t) = 1, \quad (3.9)$$

$$\chi_l^2(t) = \chi_l(t), \quad \chi_r^2(t) = \chi_r(t), \quad \chi_s^2(t) = \chi_s(t), \quad (3.10)$$

$$\chi_l(t)\chi_r(t) = 0, \quad \chi_l(t)\chi_s(t) = 0, \quad \chi_s(t)\chi_r(t) = 0, \quad (3.11)$$

we arrive at the compact expression for the output $u(t)$ of $B(\cdot)$:

$$\begin{aligned} u(t) &= (\chi_r(t) + \chi_l(t) + \chi_s(t))u(t) \\ &= \chi_r(t)m_r(v(t) - c_r) + \chi_l(t)m_r(v(t) - c_l) \\ &\quad + \chi_s(t)u_s, \end{aligned} \quad (3.12)$$

where u_s is a generic constant corresponding to the value of $u(t)$ at any active inner segment characterized by

$$\frac{u_s}{m_l} + c_l < v(t) < \frac{u_s}{m_r} + c_r. \quad (3.13)$$

Multiplying both sides of (3.5) by $\hat{\chi}_l(t)$, using (3.4), we obtain

$$\hat{\chi}_l(t)u_d(t) = \hat{\chi}_l(t)(\hat{m}_l(t)v(t) - \widehat{m_l}(t)\hat{c}_l(t)), \quad (3.14)$$

and similarly we have

$$\hat{m}_r(t)u_d(t) = \hat{\chi}_r(t)(\hat{m}_r(t)v(t) - \hat{m}_r(t)\hat{c}_r(t)). \quad (3.15)$$

Using (3.3), (3.14), (3.15) we get the following expression of $u_d(t)$:

$$\begin{aligned} u_d(t) &= (\hat{\chi}_l(t) + \hat{\chi}_r(t))u_d(t), \\ &= \hat{\chi}_l(t)(\widehat{m_l}(t)v(t) - \widehat{m_l}(t)\hat{c}_l(t)) \\ &\quad + \hat{\chi}_r(t)(\widehat{m_r}(t)v(t) - \widehat{m_r}(t)\hat{c}_r(t)). \end{aligned} \quad (3.16)$$

From (3.12), (3.16), we have the following relationship between $u(t)$ and $u_d(t)$;

$$\begin{aligned} u(t) &= u_d(t) + \hat{\chi}_r(t)(m_r(v(t) - c_r) \\ &\quad - \widehat{m_r}(t)v(t) + \widehat{m_r}(t)\hat{c}_r(t)) + \hat{\chi}_l(t) \\ &\quad \times (m_l(v(t) - c_l) - \widehat{m_l}(t)v(t) + \widehat{m_l}(t)\hat{c}_l(t)) \\ &\quad + d_0(t), \end{aligned} \quad (3.17)$$

where

$$\begin{aligned} d_0(t) &= (\chi_r(t) - \hat{\chi}_r(t))(m_r(v(t) - c_r)) \\ &\quad + (x_l(t) - \hat{\chi}_l(t))(m_l(v(t) - c_l)) \\ &\quad + \chi_s(t)u_s, \end{aligned} \quad (3.18)$$

which represents the unparametrizable part of the error between $u(t)$ and $u_d(t)$.

From (3.18), we see that $d_0(t)$ is reduced to zero if $\chi_r(t) - \hat{\chi}_r(t) = \chi_l(t) - \hat{\chi}_l(t) = \chi_s(t) = 0$. This condition is satisfied if $\widehat{m_l} = m_l$, $\widehat{m_r} = m_r$, $\hat{c}_l = c_l$, and $\hat{c}_r = c_r$, because, after initialization, the motion of $v(t)$, $u(t)$ will not be on any of the inner segments, and $u(t)$, $v(t)$ are on the upward (downward), side of $B(\cdot)$ if and only if $u_d(t)$ and $v(t)$ are on the upward (downward) side of BI$(\cdot)$.

When the parameter estimation errors are present, the above condition is not satisfied so that $d_0(t) \neq 0$ in general. However, as we show next, $d_0(t)$ is bounded.

Proposition 3.1. The unparametrizable part $d_0(t)$ of the control error $u(t) - u_d(t)$ is bounded for any $t \geq 0$.

Proof. There are three different cases to be examined:

(1) if $\chi_l(t) = 1$, $\chi_r(t) = \chi_s(t) = 0$, then

$$d_0(t) = \begin{cases} 0 & \text{for } \hat{\chi}_l(t) = 1,\ \hat{\chi}_r(t) = 0 \\ (m_l - m_r)v(t) - m_l c_l + c_r m_r & \text{for } \hat{\chi}_l(t) = 0,\ \hat{\chi}_r(t) = 1; \end{cases} \quad (3.19)$$

(2) if $\chi_r(t) = 1$, $\chi_l(t) = \chi_s(t) = 0$, then

$$d_0(t) = \begin{cases} 0 & \text{for } \hat{\chi}_l(t) = 0,\ \hat{\chi}_r(t) = 1 \\ (m_r - m_l)v(t) + m_l c_l - c_r m_r & \text{for } \hat{\chi}_l(t) = 1,\ \hat{\chi}_r(t) = 0; \end{cases} \quad (3.20)$$

(3) if $\chi_s(t) = 1$, $\chi_l(t) = \chi_r(t) = 0$, then

$$d_0(t) = \begin{cases} -m_l v(t) + c_l m_l + u_s & \text{for } \hat{\chi}_l(t) = 1,\ \hat{\chi}_r(t) = 0 \\ -m_r v(t) + c_r m_r + u_s & \text{for } \hat{\chi}_l(t) = 0,\ \hat{\chi}_r(t) = 1. \end{cases} \quad (3.21)$$

If $m_r \neq m_l$, the intersection of the two lines of $B(\cdot)$ must not be in the region of interest, by assumption. Therefore $v(t)$ is bounded and u_s is given by

$$u_s = m_l(v(t) - c_{ls}), \quad c_{ls} \in (c_l, c_r),$$
or
$$u_s = m_r(v(t) - c_{rs}), \quad c_{rs} \in (c_l, c_r). \quad (3.22)$$

This shows that in all three cases, $d_0(t)$ is bounded.

For the usual backlash characteristic, $m_r = m_l = m$, and the expression (3.22) always holds. Therefore $d_0(t)$ is bounded even if $v(t)$ is not bounded. ∇

Letting $\hat{m}_r(t) = \hat{m}_l(t) = \hat{m}(t)$ and $\widehat{mc}_r(t) = \hat{m}(t)\hat{c}_r(t)$, $\widehat{mc}_l(t) = \hat{m}(t)\hat{c}_l(t)$, we define

$$\begin{aligned} \theta_b^* &= (mc_r, m, mc_l)', \\ \theta_b(t) &= (\widehat{mc}_r(t), \hat{m}(t), \widehat{mc}_l(t))'. \end{aligned} \quad (3.23)$$

$$\begin{aligned} \phi_b(t) &= \theta_b(t) - \theta_b^*, \\ \omega_b(t) &= (\hat{\chi}_r(t), -v(t), \hat{\chi}_l(t))'. \end{aligned} \quad (3.24)$$

Finally, from (3.17), (3.23), (3.24), we obtain the parametrized expression for the backlash inverse error $u(t) - u_d(t)$ in terms of the parameter error $\phi_b(t) = \theta_b(t) - \theta_b^*$ with a bounded disturbance $d_0(t)$:

$$u(t) - u_d(t) = \phi_b^t(t)\omega_b(t) + d_0(t). \quad (3.25)$$

This parametrization holds for the discrete-time case:

$$u(k) - u_d(k) = \phi_b^t(k)\omega_b(k) + d_0(k). \quad (3.26)$$

This expression will be important for our adaptive design.

4 AN INTRODUCTORY EXAMPLE

The purpose of this section is to give an introductory example of adaptive backlash inverse. To focus on the backlash problem, we consider that the linear part of the plant in the continuous-time form is $G(D) = k_p/D$ where k_p is a known constant. For the backlash characteristic we assume that only its breakpoint parameter $c_r = -c_l = c > 0$ is unknown, while the slope $m > 0$ is known.

Our objective is to design an adaptive law to update the backlash inverse estimate and a control $u_d(t)$ to stabilize the closed-loop system and make the plant output $y(t)$ track a given reference signal $y_m(t)$ which specifies the desired system behavior.

For a discrete-time control design, the linear part of the plant is given by

$$y(t_{k+1}) = y(t_k) + k_p \int_{t_k}^{t_{k+1}} u(t)\,dt. \quad (4.1)$$

When $u(t)$ is piecewise constant and $T \triangleq t_{k+1} - t_k > 0$, then

$$y(k+1) = y(k) + Tk_p u(k). \quad (4.2)$$

Define $\bar{u}(k) = Tk_p u(k) = \bar{B}(v(k))$, where $\bar{B}(v(k)) = B(Tk_p m_l, Tk_p m_r, c_l, c_r; v(k))$ is the modified backlash characteristic with slopes $Tk_p m_l$, $Tk_p m_r$. With the non-unity gain Tk_p taken care of by the modified backlash model and $\bar{u}(k)$ renamed as $u(k)$, and from (4.2), the linear part of the plant becomes:

$$y(k+1) = y(k) + u(k). \quad (4.3)$$

In the absence of backlash our design objective would be achieved by the controller

$$u_d(k) = -y(k) + y_m(k+1). \quad (4.4)$$

In the presence of backlash we use this controller along with an adaptive scheme designed to update the backlash inverse on-line.

Since, by assumption, m is known and $c_r = -c_l = c$, we let $\hat{m}(k) = m$, $\widehat{mc}_l(k) = -\widehat{mc}_r(k) = \widehat{mc}(k)$. The backlash inverse error equation (3.26) becomes:

$$u(k) - u_d(k) = \phi(k)\omega(k) + d_0(k), \quad (4.5)$$

where $\omega(k) = \hat{\chi}_r(k) - \hat{\chi}_l(k)$, $\phi(k) = \theta(k) - \theta^*$, $\theta(k) = \widehat{mc}(k)$, and $\theta^* = mc$.

For the tracking error $e(k) = y(k) - y_m(k)$, we obtain from (4.3)–(4.5)

$$e(k) = \theta(k-1)\omega(k-1) - \theta^*\omega(k-1) + d_0(k-1), \quad (4.6)$$

and define the estimation error as:

$$\epsilon(k) = e(k) + \theta(k)\omega(k-1) - \theta(k-1)\omega(k-1). \quad (4.7)$$

This is the implementable form of $\epsilon(k)$ to be used in adaptive update laws. A simpler but unimplementable form of $\epsilon(k)$ obtained from (4.6) and (4.7) is

$$\epsilon(k) = \phi(k)\omega(k-1) + d_0(k-1). \quad (4.8)$$

Using the implementable form of $\epsilon(k)$, our update law for $\theta(k)$ based on a gradient algorithm (Goodwin and Sin, 1984; Landau, 1990) with an initial estimate $\theta(0)$ is

$$\theta(k+1) = \theta(k) - \frac{\gamma\omega(k-1)\epsilon(k)}{1 + \omega^2(k-1)} - \sigma(k)\theta(k), \quad 0 < \gamma < 1, \quad (4.9)$$

where $\sigma(k)$ is a "switching-σ signal" (Ioannou and Tsakalis, 1986). Its implementation requires *a priori* knowledge of an upper bound M on $|\theta^*|$:

$$\sigma(k) = \begin{cases} \sigma_0 & \text{for } |\theta(k)| > 2M, \\ 0 & \text{otherwise} \end{cases}, \quad 0 < \sigma_0 < \tfrac{1}{2}(1 - \gamma). \quad (4.10)$$

The stability and tracking properties of the *closed-loop system (4.3)–(4.5), (4.9) are:*

Proposition 4.1. All signals in the closed-loop system are bounded and there exist $\alpha_0 > 0$, $\beta_0 > 0$ such that

$$\sum_{k=k_1}^{k_1+k_2} e^2(k) \le \alpha_0 \sum_{k=k_1-2}^{k_1+k_2-1} d_0^2(k) + \beta_0, \quad (4.11)$$

for any $k_1 \ge 2$, $k_2 \ge 0$.

Proof. Using (4.8), (4.9) and introducing

$$\bar{\epsilon}(k) = \frac{\epsilon(k)}{\sqrt{1+\omega^2(k-1)}},$$

$$\bar{d}_0(k-1) = \frac{d_0(k-1)}{\sqrt{1+\omega^2(k-1)}}, \quad (4.12)$$

we express the time increment of $V(k) = \phi^2(k)$ as

$$V(k+1) - V(k) \le -\sigma_0\gamma\bar{\epsilon}^2(k) - \sigma_0\sigma(k)\theta^2(k) + \bar{d}_0^2(k-1). \quad (4.13)$$

This proves that $\phi(k)$ is bounded. By definition, $\omega(k)$ is bounded. Hence $e(k)$ in (4.6) is bounded, and so is $y(k)$. Finally, $u_d(k)$ in (4.4), $v(k)$ in (2.18), $u(k)$ in (2.16), and thus all closed-loop signals, are bounded.

Using (4.7), (4.9), we have that

$$e^2(k) \le 2\epsilon^2(k) + 2\omega^2(k-1)(\theta(k) - \theta(k-1))^2$$
$$\le 2\epsilon^2(k) + 2\omega^2(k-1)\left(\frac{2\gamma^2\omega^2(k-2)}{1+\omega^2(k-2)} \times \bar{\epsilon}^2(k-1) + 2\sigma^2(k-1)\theta^2(k-1)\right), \quad (4.14)$$

which, in view of (4.12), (4.13) and the boundedness of $\omega(k)$, proves (4.11). ∇

To evaluate the closed-loop system performance improvement achieved by the proposed adaptive backlash inverse, simulations were performed for the first-order plant (4.2). The plant parameters were taken as: $k_p = 3.7$ and $m_r = m_l = m = 1.3$. The parameter $c_r = -c_l = c = 1.25$ is unknown to the adaptive backlash inverse. The discrete-time time step $T = 0.1$ so the modified slope is $k_p Tm = 0.4625$.

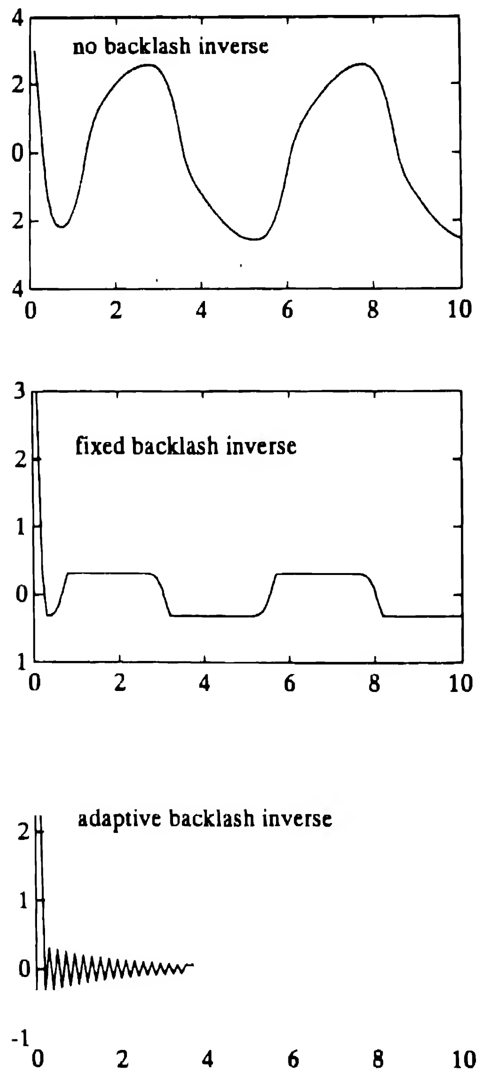


FIG. 3. Tracking errors for $y_m(t) = 10 \sin 1.26t$.

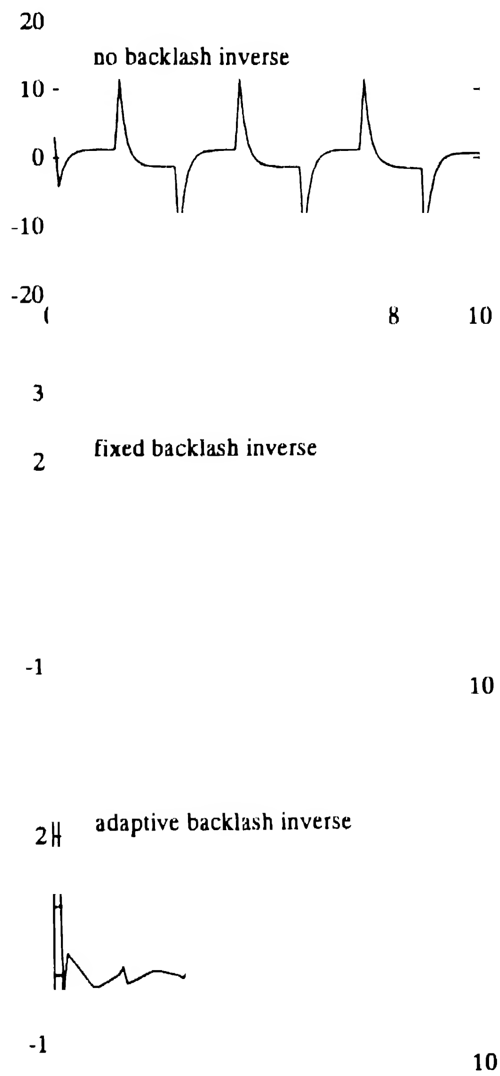


FIG. 4. Tracking errors for $y_m(t) = 10\,\mathrm{sgn}(\sin 2.2t)$ (square wave).

Three cases were studied for comparison: (1) only the controller (4.4) is applied, that is, no backlash inverse is implemented; (2) the controller (4.4) and a fixed backlash inverse (that is, the backlash inverse implemented with fixed parameter estimates) are applied; (3) the controller (4.4) and an adaptive backlash inverse are applied.

The system responses to $y_m(t) = 10 \sin 1.26t$ with $\hat{c}(0) = 1.91$ are shown in Fig. 3, and the system responses to $y_m(t) = 10\,\mathrm{sgn}(\sin 2.2t)$ (square wave) with $\hat{c}(0) = 0.59$ are shown in Fig. 4, where $\hat{c}(t) = \widehat{mc}(t)/\hat{m}(t)$ and $\hat{m}(t) = m$.

The simulation results show that the adaptive backlash inverse (case (3)) leads to major system performance improvements in all the cases of different initial conditions and different reference signals: in addition to the signal boundedness the adaptive scheme achieves convergence to zero of both tracking error and parameter error, while the control error also converges to zero because the parameter error does. The simulations results also show that a fixed backlash inverse whose parameter was either underestimated or overestimated (case (2)) is also useful; the tracking error in reduced while it is quite large in case (1) when no backlash inverse is used.

5. ADAPTIVE CONTROL DESIGN

We are now prepared to address the main problem of this paper: adaptive control design for an unknown discrete-time plant with unknown backlash at its input. Using D to denote the z-transform variable or the advance operator, as the case may be, the unknown plant to be controlled is

$$y(k) = G(D)[u](k), \quad u(k) = B(v(k)),$$
$$G(D) = \frac{Z(D)}{R(D)}. \tag{5.1}$$

Without loss of generality, the polynomials $Z(D)$ and $R(D)$ are assumed to be monic so that the high-frequency gain of $G(D)$ is one, and the actual high-frequencey gain of the plant is represented by the slope m of the backlash $B(\cdot)$

We make the following assumptions about the plant:

(A1) $G(D)$ is minimum phase;
(A2) the relative degree n^* of $G(D)$ is known;
(A3) the degree n of $R(D)$ is known;
(A4) $m \ge m_0$ for some known $m_0 > 0$, and $c_l \le 0 \le c_r$.

Our problem is to design an adaptive controller to achieve the tracking of a reference signal $y_m(k)$ by the plant output $y(k)$. To solve this problem we need a controller structure which not only achieves tracking when all the plant parameters are known but also results in a tracking error expression suitable for adaptive control design.

5.1. *Controller structure*

From the usual controller structures used in adaptive linear control we borrow the feedback part, that is, we pass the output $y(k)$ through a linear filter $\theta_y^T(a^T(D), 1)^T$, where $a(D) = (D^{-n+1}, \ldots, D^{-1})^T$, and $\theta_y \in R^n$. The new part of the controller is its forward part. It must incorporate the backlash inverse and, hence, cannot have a linear structure. In solving this new problem we still want to preserve the linear parametrization of the error equations, which will be the main tool of our adaptive design. A structure which meets this requirement is obtained by passing not only the control signal $v(k)$, but also the backlash inverse signals $\chi_r(k)$ and $\chi_l(k)$ through the filter formed of $a(D)$ and adjustable parameters θ_v, θ_r, $\theta_l \in R^{n-1}$. Introducing the four regressors:

$$\omega_r(k) = a(D)[\chi_r](k), \quad \omega_l(k) = a(D)[\chi_l](k), \tag{5.2}$$

$$\omega_v(k) = a(D)[v](k), \quad \omega_y(k) = (a^T(D), 1)^T[y](k), \tag{5.3}$$

which multiply the corresponding adjustable parameters, we propose the following linear-like structure of the nonlinear adaptive controller:

$$u_d(k) = \theta_r^T\omega_r(k) + \theta_l^T\omega_l(k) + \theta_v^T\omega_v(k) + \theta_y^T\omega_y(k) + y_m(k+n^*). \tag{5.4}$$

This controller is shown in Fig. 5 where $\chi_r(t)$ and $\chi_l(t)$ are obtained from the logic block L which implements (3.1) and (3.2).

The output $u_d(k)$ of this controller is applied to the adaptive backlash inverse in order to generate the plant control input:

$$v(k) = \widehat{BI}(u_d(k)). \tag{5.5}$$

We now show that for this controller there exist matched values of the adjustable parameters resulting in exact tracking with internal stability.

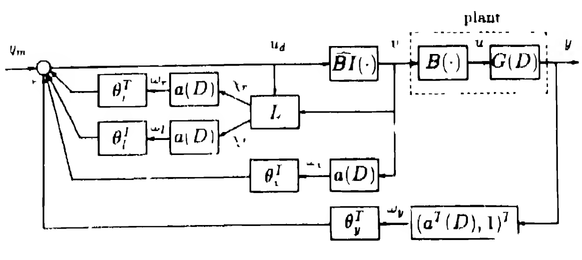


FIG. 5. The adaptive inverse controller structure.

Lemma 5.1. There exist matched values θ_r^*, θ_l^*, θ_v^*, θ_y^*, θ_b^* of θ_r, θ_l, θ_v, θ_b, and θ_y with which the controller (5.4) achieves the closed-loop global stability and tracking $y(k+n^*) = y_m(k+n^*)$.

Proof. We first express the matched values of the parameters θ_r, θ_l and θ_v in the forward part of the controller in terms of the backlash parameters $\theta_b^* = (mc_r, m, mc_l)^T$ multiplied by a parameter $\theta_u^* \in R^{n-1}$.

$$\theta_r^* = -\theta_u^* mc_r, \quad \theta_l^* = -\theta_u^* mc_l, \quad \theta_v^* = -\theta_u^* m. \tag{5.6}$$

When then define θ_u^* and the matched value θ_y^* of the feedback parameter θ_y as the solution of the Diophantine equation:

$$\theta_u^{*T} a(D) R(D) + \theta_y^{*T}(a^T(D), 1)^T Z(D) = R(D) - Z(D) D^{n^*}. \tag{5.7}$$

Substituting the matched values in the controller (5.4) we see that its forward part is a linear parametrization of the nonlinear term $\theta_u^{*T} a(D)[-\theta_b^* \omega_b](k)$, namely:

$$\theta_u^{*T} a(D)[-\theta_b^* \omega_b](k) = \theta_r^{*T}\omega_r(k) + \theta_l^{*T}\omega_l(k) + \theta_v^{*T}\omega_v(k). \tag{5.8}$$

With the matched values and (5.8) the controller (5.4) has the form:

$$u_d(k) = \theta_u^* a(D)[-\theta_b^* \omega_b](k) + \theta_y^{*T}(a^T(D), 1)^T[y](k) + y_m(k+n^*). \tag{5.9}$$

The right inverse BI(·) of the backlash B(·) is also matched, and hence

$$u(k) = \mathrm{B}(\mathrm{BI}(u_d(k))) = u_d(k) = -\theta_b^* \omega_b(k). \tag{5.10}$$

On the other hand, when both sides of (5.7) are divided by $R(D)$ and then operated on $u(k)$, the resulting identity is

$$u(k) = \theta_u^{*T} a(D)[u](k) + \theta_y^{*T}(a^T(D), 1)^T[y](k) + y(k+n^*). \tag{5.11}$$

Equating $u(k)$ of (5.11) with $u_d(k)$ of (5.9) and using (5.10) prove $y(k+n^*) = y_m(k+n^*)$. The closed-loop system is globally stable because with the matched values the closed-loop poles are zeros of $D^{n+n^*-1}Z(D)$. ∇

It is due to the identity (5.8) that the new controller structure (5.4) will lead to a convenient linear form of the tracking error expression. For a more compact notation, we let

$$\theta(k) = (\theta_r^T(k), \theta_l^T(k), \theta_v^T(k), \theta_y^T(k), \theta_b^T(k))^T,$$
$$\theta^* = (\theta_r^{*T}, \theta_l^{*T}, \theta_y^{*T}, \theta_v^{*T}, \theta_b^{*T})^T, \tag{5.12}$$
$$\omega(k) = (\omega_r^T(k), \omega_l^T(k), \omega_v^T(k), \omega_y^T(k), \omega_b^T(k))^T. \tag{5.13}$$

Lemma 5.2. The controller (5.4) with arbitrary $\theta_r(k)$, $\theta_l(k)$, $\theta_v(k)$, $\theta_y(k)$, $\theta_b(k)$ results in the tracking error consisting of a linear part $\phi^T(k)\omega(k)$ and a bounded part $d_1(k)$, that is:

$$e(k) \triangleq y(k) - y_m(k) = \phi^T(k-n^*)\omega(k-n^*) + d_1(k), \ \phi(k) = \theta(k) - \theta^*, \tag{5.14}$$

$$d_1(k) = d_0(k-n^*) - \theta_u^{*T} a(D)[d_0](k-n^*). \tag{5.15}$$

Proof. Recall from (3.16), (3.23) and (3.24) that the expression of the backlash inverse estimate $\widehat{\mathrm{BI}}(\cdot)$ is $u_d(k) = -\theta_b^T(k)\omega_b(k)$ so that (3.26) gives

$$u(k) = -\theta_b^{*T}\omega_b(k) + d_0(k). \tag{5.16}$$

Substituting (5.16) in the right side of (5.11) and (3.26) in the left side of (5.11), and using (5.8), we obtain

$$u_d(k) + \phi_b^T(k)\omega_b(k) + d_0(k) = \theta_y^{*T}\omega_y(k) + y_m(k+n^*) + e(k+n^*) + \theta_r^{*T}\omega_r(k) + \theta_l^{*T}\omega_l(k) + \theta_v^{*T}\omega_v(k) + \theta_u^{*T}a(D)[d_0](k). \tag{5.17}$$

Finally, we substitute (5.17) into (5.4) to get (5.14). ∇

We have thus obtained a tracking error equation with a linear parametrization and an unknown, but bounded, disturbance.

5.2. *Adaptive scheme*

Our major task now is to design an adaptive scheme to update the parameters of the backlash inverse (2.21) and the controller (5.4) to guarantee the signal boundedness for the closed-loop system. This task is achievable with the tools of adaptive linear control (Ioannou and Tsakalis, 1986; Egardt, 1979; Kreisselmeier and Anderson, 1986; Narendra and Annaswamy, 1989; Middleton *et al.*, 1988; Praly, 1990).

One update law for $\theta(k)$ with an initial estimate $\theta(0)$ suggested by the form of the tracking error equation (5.14) is

$$\theta(k+1) = \theta(k) - \frac{\gamma\omega(k-n^*)\epsilon(k)}{1+\omega^T(k-n^*)\omega(k-n^*)} - \sigma(k)\theta(k), \quad 0 < \gamma < 1, \tag{5.18}$$

where $\epsilon(k)$ is the estimation error:

$$\epsilon(k) = e(k) + (\theta(k) - \theta(k-n^*))^T\omega(k-n^*), \tag{5.19}$$

and $\sigma(k)$ is a "switching-σ modification" (Ioannou and Tsakalis, 1986) whose implementation requires *a priori* knowledge of an upper

bound M on the Euclidean norm $\|\theta^*\|_2$ of θ^*:

$$\sigma(k) = \begin{cases} \sigma_0 & \text{for } \|\theta(k)\|_2 > 2M \\ 0 & \text{otherwise} \end{cases}, \quad 0 < \sigma_0 < \tfrac{1}{2}(1-\gamma). \tag{5.20}$$

Although not shown in (5.18) we use projection to ensure that $\widehat{m}(k) \geq m_0$ and $\widehat{mc_l}(k) \leq 0 \leq \widehat{mc_r}(k)$ to implement the adaptive backlash inverse (2.21).

This adaptive control scheme has the following stability and tracking properties.

Theorem 5.1. All signals in the closed-loop system are bounded and there exist $\alpha_0 > 0$, $\beta_0 > 0$ such that

$$\sum_{k=k_1}^{k_1+k_2} e^2(k) \leq \alpha_0 \sum_{k=k_1-n_0}^{k_1+k_2-n^*} d_0^2(k) + \beta_0, \tag{5.21}$$

for $n_0 = 2n^* + n - 1$ and any $k_1 \geq n_0$, $k_2 \geq 0$.

Proof. The first part of the proof, which shows the boundedness of the update law, is standard. Substituting (5.14) in (5.19) results in

$$\epsilon(k) = \phi(k)\omega(k-n^*) + d_1(k). \tag{5.22}$$

Using (5.18), (5.22) and introducing

$$\bar{\epsilon}(k) = \frac{\epsilon(k)}{\sqrt{1+\omega^T(k-n^*)\omega(k-n^*)}},$$

$$\bar{d}_1(k) = \frac{d_1(k)}{\sqrt{1+\omega^T(k-n^*)\omega(k-n^*)}}, \tag{5.23}$$

we express the time increment of $V(k) = \phi^T(k)\phi(k)$ as

$$V(k+1) - V(k) \leq -\sigma_0\gamma\bar{\epsilon}^2(k) - \sigma_0\sigma(k)\theta^T(k)\theta(k) + \bar{d}_1^2(k). \tag{5.24}$$

This proves that $\phi(k) = \theta(k) - \theta^*$ is bounded. In view of (5.22), (5.23) and the boundedness of $d_1(k)$, this implies that $\bar{\epsilon}(k)$ is also bounded.

In the second part of the proof we use a novel technique to show the closed-loop signal boundedness. Let us introduce

$$\bar{\omega}(k) = (\omega_u^T(k), \omega_y^T(k))^T, \quad \omega_u(k) = a(D)[u](k). \tag{5.25}$$

It can be shown that there exist bounded sequences $F_1(k) \in R^{4n\times(2n-1)}$, $g_1(k) \in R^{4n}$, $g_2(k) \in R^{2n-1}$ and constant $F_2 \in R^{(2n-1)\times 4n}$ such that

$$\omega(k) = F_1(k)\bar{\omega}(k) + g_1(k),$$
$$\bar{\omega}(k) = F_2\omega(k) + g_2(k). \tag{5.26}$$

Then we use (5.7) and the fact that $y(k) = (Z(D)/R(D))[u](k)$ to obtain

$$\theta_u^{*T}\omega_u(k) + \theta_y^{*T}\omega_y(k) = u(k) - y(k+n^*). \tag{5.27}$$

Using this equality and the definition of $\bar{\omega}(k)$, we express

$$\bar{\omega}(k+1) = A^*\bar{\omega}(k) + b^*y(k+n^*) = A^*\bar{\omega}(k) + b^*\frac{Z(D)}{R(D)}[u](k+n^*), \tag{5.28}$$

for some constant matrix $A^* \in R^{(2n-1)\times(2n-1)}$ and constant vector $b^* \in R^{2n-1}$. Since the first component of $\bar{\omega}(k)$ is $u(k-n+1)$, it follows that for $c^* = (1, 0, \ldots, 0)^T \in R^{2n-1}$,

$$D^{-n+1}u(D) = c^{*T}(DI - A^*)^{-1}b^*\frac{Z(D)}{R(D)}D^{n^*}u(D), \tag{5.29}$$

which implies that

$$c^*(DI - A^*)^{-1}b^* = \frac{R(D)}{Z(D)D^{n+n^*-1}},$$
$$\det(DI - A^*) = D^{n+n^*-1}Z(D), \tag{5.30}$$

that is, A^* is a stable matrix. Using (5.19) and the first equality of (5.28), we have

$$\bar{\omega}(k+1) = A^*\bar{\omega}(k) + b^*(y_m(k+n^*) + \epsilon(k+n^*) - (\theta(k+n^*) - \theta(k))^T\omega(k)). \tag{5.31}$$

From (5.30), all eigenvalues of A^* are inside the unit circle of the complex plane. Therefore there exists a non-singular constant matrix $Q \in R^{(2n-1)\times(2n-1)}$ such that $\|QA^*Q^{-1}\|_2 < 1$. Define the vector norm $\|\cdot\|$ in R^{2n-1} by: $\|x\| = \|Qx\|_2$. From (5.26), there exist constants $c_i > 0$, $\bar{c}_i > 0$, $i = 1, 2$, such that for all $k \geq 0$

$$\|\bar{\omega}(k)\| \leq c_1\|\omega(k)\|_2 + c_2, \tag{5.32}$$

$$\|\omega(k)\|_2 \leq \bar{c}_1\|\bar{\omega}(k)\| + \bar{c}_2. \tag{5.33}$$

Introducing $x(k) = |\bar{\epsilon}(k+n^*)| + \|\theta(k+n^*) - \theta(k)\|_2$, using (5.18), (5.24) and (5.32) it can be shown that ther exist constants $c_3 > 0$, $c_4 > 0$ such that

$$\sum_{k=k_1}^{k_1+k_2} x^2(k) \leq c_3 \sum_{k=k_1-n^*}^{k_1+k_2} \frac{1}{1+\|\bar{\omega}(k)\|^2} + c_4. \tag{5.34}$$

It follows from (5.30), (5.31), (5.33), there exist constants $a_0 \in (0, 1)$, $c_5 > 0$, $c_6 > 0$ such that

$$\|\bar{\omega}(k+1)\| \leq (a_0 + c_5x(k))\|\bar{\omega}(k)\| + c_6. \tag{5.35}$$

Substituting (5.19) in (5.31) and using (5.14), (5.26), we obtain

$$\bar{\omega}(k+1) = (A^* + b^*\phi^T(k)F_1(k))\bar{\omega}(k) + g_3(k), \tag{5.36}$$

for some bounded sequence $g_3(k) \in R^{2n-1}$. Since $\phi(k)$, $F_1(k)$ and $g_3(k)$ in (5.36) are bounded, $\bar{\omega}(k)$ grows at the most exponentially, that is, there exist constants $c_7 > 0$, $c_8 > 0$ such that for all $k \geq 0$

$$\|\bar{\omega}(k+1)\| \leq c_7 \|\bar{\omega}(k)\| + c_8. \quad (5.37)$$

Now we show the boundedness of $\bar{\omega}(k)$ by contradiction. Assume that $\bar{\omega}(k)$ grows unboundedly. Then, in view of (5.37), given any $\delta_0 > 0$ and $k_2 > 0$, we can find $\delta \in (0, \delta_0]$ and $k_1 > 0$ such that

$$\|\bar{\omega}(k)\| \geq \frac{1}{\delta_0}, \quad k \in \{k_1 - n^*, \ldots, k_1 - 1\}, \quad (5.38)$$

$$\|\bar{\omega}(k)\| = \frac{1}{\delta}, \quad k = k_1, \quad (5.39)$$

$$\|\bar{\omega}(k)\| \geq \frac{1}{\delta}, \quad k \in \{k_1 + 1, \ldots, k_1 + k_2 + 1\}. \quad (5.40)$$

Therefore, for $j \in \{0, \ldots, k_2\}$, the state transition function of $\|\bar{\omega}(k+1)\| = (a_0 + c_5 x(k)) \|\bar{\omega}(k)\|$ satisfies

$$\begin{aligned}
\prod_{k=k_1}^{k_1+j} (a_0 + c_5 x(k)) &\leq \left(a_0 + \frac{c_5}{j+1} \sum_{k=k_1}^{k_1+j} x(k)\right)^{j+1} \\
&\leq \left(a_0 + (c_5\sqrt{c_3(n^*+1)})\,\delta_0 + c_5 \frac{\sqrt{c_4}}{\sqrt{j+1}}\right) \\
&\leq (a_0 + (c_5\sqrt{c_3(n^*+1)})\,\delta_0)^{j+1} \\
&\quad \times \left(\frac{c_5\sqrt{c_4}}{a_0 + c_5\sqrt{c_3(n^*+1)}}\right)^{\sqrt{j+1}}. \quad (5.41)
\end{aligned}$$

With any δ_0 satisfying $a_0 + (c_5\sqrt{c_3(n^*+1)})\,\delta_0 < 1$, (5.35) and (5.41) imply that

$$\|\bar{\omega}(k_1 + j + 1)\| \leq \prod_{k=k_1}^{k_1+j} (a_0 + c_5 x(k)) \|\bar{\omega}(k_1)\| + c_9, \quad (5.42)$$

for some constant $c_9 > 0$, and $\prod_{k=k}^{k_1+j} (a_0 + c_5 x(k)) < \frac{1}{2}$ for any $j \geq j_1$ and some $j_1 \geq 0$. Hence, for

$$\delta_0 \in \left(0, \min\left\{\frac{1-a_0}{c_5\sqrt{c_3(n^*+1)}}, \frac{1}{2c_9}\right\}\right) \quad \text{and} \quad k_2 \geq j_1,$$

(5.42) implies that $\|\bar{\omega}(k_1 + j + 1)\| < 1/\delta$ for any $j \in \{j_1, \ldots, k_2\}$, which is a contradiction. Hence $\bar{\omega}(k)$ is bounded, and so is $\omega(k)$.

Next we use this fact to prove the bound given by (5.21). From (5.18) and (5.19), we first evaluate a bound on the sum of $e^2(k)$ in terms of $\epsilon^2(k)$ and $\sigma^2(k-1)\theta^T(k-1)\theta(k-1)$ and their past values. Since this sum contains $d_1(k)$ and its past values, we use (5.15) to express $d_1(k)$ in terms of $d_0(k - n^*)$ and its past values and finally obtain (5.21). ∇

We have thus shown that the adaptive law (5.18) ensures robustness with respect to the bounded disturbance $d_0(k)$. Another advantage of this update law is that the asymptotic tracking is achieved if the adaptive backlash inverse converges to the exact one, that is when $d_0(k)$ disappears for large k, see (5.21).

We have not yet shown that, in general, the tracking error $e(k)$ converges to zero. The dependence of $d_0(k)$ on the parameter error suggests that this will be so if the adaptive system has sufficiently rich signals. Extensive adaptive backlash inverse simulations showed that with rich signals the backlash inverse parameters converge to their matched values so that the asymptotic tracking is achieved.

6. CONCLUSIONS

This paper has presented what appear to be the first formulation and solution of the adaptive control problem for systems with backlash. To achieve this, we first demonstrated the right invertibility of a general backlash model and parametrized the error expression of a backlash inverse estimate needed for continuous and discrete-time implementation. We then introduced a new linear-like structure for a nonlinear controller capable of cancelling the effects of backlash. This controller structure makes it possible to obtain a linear error equation, with the effect of an inaccurate backlash inverse represented by a bounded disturbance. From this point on, a robust adaptive update law was designed to guarantee global signal boundedness. Simulation results showed major system performance improvements.

Acknowledgements—This work was supported by National Science Foundation under Grant ECS 87-15811, by the Air Force Office of Scientific Research under Grant AFOSR 90-0011 and by a Ford Motor Company grant. We are thankful to Jim Winkleman and Doug Rhode of Ford Motor Company for stimulating this research and to Darrel Recker of University of Illinois for many helpful discussions

REFERENCES

Egardt, B (1979). *Stability of Adaptive Controllers* Springer-Verlag, Berlin

Goodwin, G C. and K S. Sin (1984) *Adaptive Filtering Prediction and Control.* Prentice-Hall, Englewood Cliffs, NJ.

Ioannou, P. A and K. Tsakalis (1986). Robust discrete time adaptive control In K. S. Narendra (Ed), *Adaptive and Learning Systems: Theory and Applications.* Plenum Press, New York.

Krasnosel'skii, M. A. and A. V. Pokrovskii (1983) *Systems with Hysteresis* Springer-Verlag, Berlin.

Kreisselmeier, G. and B. D. O Anderson (1986) Robust

model reference adaptive control *IEEE Trans Aut Control*, **31**, 127–133
Landau, I D (1990) *System Identification and Control Design* Prentice-Hall, Englewood Cliffs NJ
Mayergoyz, I D (1991) *Mathematical Models of Hysteresis* Springer-Verlag, Berlin
Middleton, R H , G C Goodwin, D J Hill and D Q Mayne (1988) Design issues in adaptive control *IEEE Trans Control*, **33**, 50–58
Narendra, K S and A M Annaswamy (1989) *Stable Adaptive Systems* Prentice-Hall, Englewood Cliffs, NJ
Netushil, A (1973) *Theory of Automatic Control* Mir Moscow
Praly, L (1990) Almost exact modelling assumption in adaptive control *Int J Control*, **51**, 643–668
Recker, D A , P V Kokotović, D S Rhode and J R Winkelman (1991) Adaptive nonlinear control of systems containing a dead-zone *Proc of the 30th IEEE Conference on Decision and Control*, pp 2111–2115, Brighton U K
Tao, G and P V Kokotović (1991) Adaptive control of plants with unknown dead-zones Tech Report CCEC-91 1006, UCSB *Proc of the* 1992 *American Control Conference* Chicago, IL, pp 2710–2714
Truxal, J G (1958) *Control Engineers' Handbook* McGraw-Hill, New York

Automatica Vol 29, No 2 pp 337–349 1993
Printed in Great Britain

0005 1098/93 $6 00 + 0 00

Performance Limitations of Non-minimum Phase Systems in the Servomechanism Problem*†

L. QIU‡ and E. J DAVISON§

A fundamental limitation exists in the achievable transient performance which is possible to be obtained in the tracking and disturbance rejection of a non-minimum phase system, and this limitation can be characterized completely by the number and locations of the right-half plane zeros.

Key Words—Linear optimal regulator, servomechanisms, non-minimum phase systems, transmission zeros, cheap control

Abstract—This paper studies the cheap regulator problem and the cheap servomechanism problem for systems which may be non-minimum phase The study extends some well-known properties of "perfect regulation" and the 'perfect tracking and disturbance rejection of minimum phase systems to non-minimum phase systems It is shown that perfect rejection to disturbances applied to the plant input can be achieved no matter whether the system is minimum phase or non minimum phase, whereas a fundamental limitation exists in the achievable transient performance of tracking and rejection to disturbances applied to the plant output for a non-minimum phase system, and that this limitation can be simply and completely characterized by the number and locations of those zeros of the system which lie in the right half of the complex plane Furthermore, this limitation provides a quantitative measure of the "degree of difficulty" which is inherent in the control of such non-minimum phase systems

1 INTRODUCTION

It has long been realized that minimum phase systems have certain advantages over non-minimum phase systems, for example, right-invertible minimum phase systems can achieve perfect regulation (Kwakernaak and Sivan, 1972b; Francis, 1979; Scherzinger and Davison, 1985) and perfect tracking/disturbance rejection (Davison and Chow, 1977; Davison and Scherzinger, 1987). These properties, however, are not possessed by non-minimum phase systems. In addition, a non-minimum phase system, unlike minimum phase systems, has various fundamental limitations associated with the achievable closed loop transfer matrix (Cheng and Desoer, 1980) the achievable closed loop gain margin (Tannenbaum, 1980) LQG loop transfer recovery (Stein and Athans, 1987; Zhang and Freudenberg, 1990) sensitivity or complementary sensitivity minimization (Freudenberg and Looze, 1985; Francis, 1987) model reference adaptive control (Miller and Davison, 1989) etc. On the other hand, it has been recognized that not all non-minimum phase systems behave in the same way; for example, some non-minimum phase systems produce results which are "almost as good" as minimum phase systems, whereas other non-minimum phase systems are indeed "almost impossible" to control. It is the purpose of this paper to study the quality of non-minimum phase systems with respect to tracking and disturbance rejection. We will show that for each non-minimum phase system there exists a fundamental limitation on the achievable transient response of the system. This limitation has a simple characterization in terms of the number and the locations of those zeros of the system which lie in the open-right complex plane, and provides a quantitative measure of the "degree of difficulty" which is inherent in the control of a non-minimum phase system

In this paper, we first study the cheap linear quadratic regulator (LQR) problem for non-minimum phase systems. Consider a linear

* Received 15 October 1991, revised 17 June 1992, received in final form 2 July 1992 The original version of this paper was presented at the 11th IFAC World Congress on Automatic Control at the Service of Mankind which was held in Tallinn, Estonia during August, 1990 The Published Proceedings of this IFAC Meeting may be ordered from Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U K This paper was recommended for publication in revised form by Associate Editor P Dorato under the direction of Editor H Kwakernaak

† This work has been supported by the Natural Sciences and Engineering Research Council of Canada under grant no A4396

‡ The Institute for Mathematics and its Applications, University of Minnesota 514 Vincent Hall, 206 Church Street S E , Minneapolis, MN, U S A

§ Department of Electrical Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4

time-invariant system described by

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0, \\ y &= Cx + Du, \end{aligned} \tag{1}$$

where u, x, y are finite dimensional vectors depending on the time t. Assume that (A, B, C, D) is stabilizable and detectable, and consider also the associated optimal cost functional

$$J_\epsilon = \min_u \int_0^\infty (y'y + \epsilon^2 u'u)\,dt. \tag{2}$$

The cheap LQR problem concerns the limit $J_0 := \lim_{\epsilon\to 0} J_\epsilon$. It is shown in this paper that J_0 has a very simple expression when (A, B, C, D) has a particular special structure. This expression is not very important *per se* since very few systems have such a special structure. However, the significance of the result lies in the fact that for every transfer matrix, there always exists a realization which has such a special structure. Therefore the result has important applications in control problems which depend only on the system transfer matrix rather than on the internal realization of the system. The cheap optimal servomechanism problem, which is the focus of this paper, belongs to this class of problems. An expression for J_0 involving the concept of "zero directions" was obtained by Shaked (Grimble and Johnson, 1988) for the case when D is a zero matrix and A, B, C are general matrices; this expression, although interesting, is quite complicated and is not convenient to use in our application.

Consider now a system with noise corrupted input and output:

$$\begin{aligned} \dot{x} &= Ax + B(u + \xi), \quad x(0) = 0, \\ y &= Cx + D(u + \xi) + \eta, \end{aligned} \tag{3}$$

where ξ is the input disturbance, η is the output disturbance. We assume again that (A, B, C, D) is stabilizable and detectable. A control problem which often arises is to design a controller for system (3) such that the overall system is internally stable and such that the output y asymptotically tracks a reference signal y_{ref} for arbitrary ξ, η and y_{ref} contained in a certain class of signals. This problem is called a *servomechanism problem*. It is well-known that a general servomechanism problem can be treated as one with the reference signal being equal to zero since the distinguishing role between y_{ref} and η disappears if the tracking error $y - y_{\text{ref}}$ is taken to be the output under consideration. Therefore, we will assume that $y_{\text{ref}} = 0$ throughout the paper. In practice, a controller which solves the servomechanism problem is also required to have a good transient response, i.e. it is desired that the closed loop system should have a "fast speed of response" without "excessive peaking/oscillation" occurring in the output y and other system variables, as they approach their steady state values. To achieve such a response, we seek a controller which generates an input to achieve the following optimal quadratic cost functional

$$J_\epsilon = \min_u \int_0^\infty (y'y + \epsilon^2 \bar{u}'\bar{u})\,dt, \tag{4}$$

where $\bar{u}$ is a variable associated with the transient behaviour of the input u. The problem concerning the limit $J_0 := \lim_{\epsilon\to 0} J_\epsilon$ is called the *cheap servomechanism problem*. It is clear that J_ϵ and J_0 depend on the controller structure used since the controller structure will determine the physical meaning of $\bar{u}$. In this paper, we first look at the ideal case: we assume that the system parameters are exactly known and that the disturbances are measurable. In this case, a feedforward controller structure can be used (Davison, 1973). It is shown in this case that J_0 is a quadratic form on the output disturbance only, and that a norm of this quadratic form can be characterized explicitly by the locations of the zeros of system (3) which lie in the open right hand side of the complex plane. In practice, the system parameters are always somewhat uncertain and it is often impossible to measure the disturbances; in this case, the servomechanism problem is still solvable but the controller has to contain an internal model of the disturbances, see e.g. Francis and Wonham (1976) and Davison (1976). Such a controller is called a robust servomechanism controller. We will show in this case that J_0 has the same characteristics as in the feedforward controller case, i.e. J_0 is a quadratic form on the output disturbance only and a norm of this quadratic form can be characterized explicitly in the same way by the locations of those zeros of sytem (3) which lie in the open right hand side of the complex plane. The significance of the results for the feedforward controller lies in that it tells us what is the best possible result that can be achieved in the ideal case, while the results for the robust servomechanism controller show that even though the controller does not have as much information, the limiting performance is identical to the ideal case.

The structure of this paper is as follows: Section 2 gives some preliminary material on the factorization of a system into the product of an inner system and a minimum phase systems.

Section 3 gives the main result on the cheap LQR problem for a special system resulting from the factorization given in Section 2. Section 4 studies the servomechanism problem with cheap quadratic cost using feedforward controllers. Section 5 studies the same problem using robust controllers. It is assumed in Sections 4 and 5 that the disturbances are constant signals. Section 6 extends the result obtained in Sections 4 and 5 to sinusoidal disturbances. Section 7 contains an example. Section 8 contains conclusions.

2 PRELIMINARIES

The transfer matrix F of the system (1) or (3) is given by

$$F(s) = D + C(sI - A)^{-1}B. \tag{5}$$

Conversely, a four-tuple of real matrices (A, B, C, D) is said to be a realization of a proper real rational matrix (transfer matrix in short) F if (5) is satisfied. Throughout this paper, the following notation is used to divide the complex plane into three parts: $\mathbb{C}^+ = \{s \in \mathbb{C} : \mathrm{Re}(s) > 0\}$, $\mathbb{C}^0 = \{s \in \mathbb{C} : \mathrm{Re}(s) = 0\}$, $\mathbb{C}^- = \{s \in \mathbb{C} : \mathrm{Re}(s) < 0\}$. A transfer matrix is said to be stable if all of its poles are contained in $\mathbb{C}^-$, and a square constant real matrix is said to be stable if all of its eigenvalues are contained in $\mathbb{C}^-$.

The zeros of a transfer matrix are defined to be the roots of the numerator polynomials of the nonzero elements of its Smith–McMillan form. A transfer matrix is said to be *minimum phase* if it has no zeros in $\mathbb{C}^+$; otherwise it is said to be *non-minimum phase*. The zeros of system (1), system (3) or simply a realization (A, B, C, D) are defined to be the roots of the invariant polynomials of the matrix $\begin{bmatrix} A - \lambda I & B \\ C & D \end{bmatrix}$. Similarly, we can define the concepts of minimum phase and non-minimum phase for system (1), system (3) and a four-tuple (A, B, C, D), as was done for a transfer matrix. Although the zeros of a real rational matrix and those of its realization may be different, their minimum phase or non-minimum phase property is always the same as long as the realization is stabilizable and detectable.

Associated with any realization (A, B, C, D) in which A is stable, there are two Lyapunov equations:

$$AP + PA' = -BB', \tag{6}$$

$$A'Q + QA = -C'C. \tag{7}$$

The solutions P, Q to equations (6)–(7) are called the controllability grammian and the observability grammian of (A, B, C, D), respectively. A minimal realization (A, B, C, D) of a stable transfer matrix F is called a *balanced realization* if the solutions P, Q to equations (6)–(7) are diagonal and equal. It is shown in Moore (1981) that every stable transfer matrix has a balanced realization. Procedures to find a balanced realization from any minimal realization of a stable transfer matrix are given in Moore (1981) and Laub *et al.* (1987).

A stable transfer matrix F is called *inner* if $F'(-s)F(s) = I$. All the zeros of an inner matrix must be located in $\mathbb{C}^+$.

Lemma 1 (Glover, 1984). Let (A, B, C, D) be a balanced realization of an inner matrix F and let P, Q be the solutions to equations (6)–(7). Then

(a) $P = Q = I$,

(b) $D'D = I$,

(c) $D'C + B' = 0$, $DB' + C = 0$.

A transfer matrix F is said to be *right-invertible* if F has full row rank for at least one $s \in \mathbb{C}$. If (A, B, C, D) is any realization of F, then the right-invertibility of F is equivalent to the fact that $\begin{bmatrix} A - \lambda I & B \\ C & D \end{bmatrix}$ has full row rank for at least one $\lambda \in \mathbb{C}$. Therefore, no confusion will be caused when we talk about the right-invertibility of system (1), system (3) or realization (A, B, C, D).

The following factorization result serves as a foundation for our development. It is noted here that when the poles and/or zeros of two transfer matrices are compared in the following, we consider not only their values but also their multiplicities in the Smith–McMillan sense.

Lemma 2. A transfer matrix F can always be factorized as $F = F_1F_2$ such that F_1 is inner, F_2 is minimum phase and right-invertible, and the unstable poles of F_2 are equal to the unstable poles of F.

The result has been known for a long time but its original proof is hard to trace. Readers are referred to Qiu and Davison (1990) for a proof and Zhang and Freudenberg (1990) for a proof of its dual version.

Given a transfer matrix F, let $F = F_1F_2$ be the factorization described in Lemma 2. Let (A_1, B_1, C_1, D_1) be a balanced realization of F_1 and let (A_2, B_2, C_2, D_2) be any stabilizable and detectable realization of F_2. Then a stabilizable and detectable realization of F is given by

$$A = \begin{bmatrix} A_1 & B_1C_2 \\ 0 & A_2 \end{bmatrix}, \quad B = \begin{bmatrix} B_1D_2 \\ B_2 \end{bmatrix}, \tag{8}$$
$$C = [C_1 \quad D_1C_2], \quad D = D_1D_2.$$

This realization is called a *factorized realization* of F and plays an important role in the development.

The following lemma gives some useful properties of right-invertible transfer matrices with respect to the factorization described in Lemma 2.

Lemma 3. Let F be a right-invertible transfer matrix and let $F = F_1F_2$ be the factorization described in Lemma 2. Then F_1 is square, the zeros of F_1 are equal to those zeros of F contained in $\mathbb{C}^+$, and the poles of F_1 are equal to the negatives of the zeros of F_1.

Proof. Let F be $r \times m$. Then F_1 has r rows and at most r columns. If F_1 has less than r columns, then the rank of F is less than r for all $s \in \mathbb{C}$, which contradicts the assumption that F is right-invertible. Therefore F_1 must have r columns, i.e. it must be square. It follows directly from the identity $F_1^{-1}(s) = F_1'(-s)$ that the poles of F_1 are equal to the negatives of the zeros of F_1.

Let (A, B, C, D) be a factorized realization of F. Those zeros of F which are contained in $\mathbb{C}^+$ are the complex numbers $\lambda \in \mathbb{C}^+$ which make the matrix

$$\begin{bmatrix} A_1 - \lambda I & B_1C_2 & B_1D_2 \\ 0 & A_2 - \lambda I & B_2 \\ C_1 & D_1C_2 & D_1D_2 \end{bmatrix}$$

reduce rank. Adding $-B_1D_1^{-1}$ times of the third row to the first row and multiplying the third row by D_1^{-1}, we transform the above matrix into

$$\begin{bmatrix} A_1 - B_1D_1^{-1}C_1 - \lambda I & 0 & 0 \\ 0 & A_2 - \lambda I & B_2 \\ D_1^{-1}C_1 & C_2 & D_2 \end{bmatrix}.$$

Hence the zeros of F in $\mathbb{C}^+$ are the eigenvalues of $A_1 - B_1D_1^{-1}C_1$. Notice from Lemma 1 that

$$\begin{aligned} A_1 - B_1D_1^{-1}C_1 &= A_1 + B_1D_1^{-1}D_1B_1' \\ &= A_1 + B_1B_1' \\ &= A_1 - (A_1 + A_1') = -A_1'. \end{aligned}$$

This completes the proof. □

We end this section with a few words about the norm of quadratic forms. A (real) quadratic form f on $v \in \mathbb{R}^p$ is a function of the form $f(v) = v'Qv$ for some symmetric matrix $Q \in \mathbb{R}^{p\times p}$. Therefore a norm of Q gives a norm of the quadratic form f. The trace norm of Q, i.e. the sum of the singular values of Q, is of particular interest. If Q is positive semi-definite, then its trace norm is equal to its trace. The following lemma gives a physical meaning to the trace norm. Let $\mathscr{E}(\cdot)$ denote the expectation operator.

Lemma 4 (Levine and Athans, 1970). Let $Q \in \mathbb{R}^{p\times p}$ be a positive semi-definite matrix and let v be a random vector in $\mathbb{R}^p$ with $\mathscr{E}(v) = 0$ and $\mathscr{E}(vv') = I$. Then $\mathscr{E}(v'Qv) = \operatorname{tr} Q$.

3. CHEAP LQR PROBLEM

Consider the cheap LQR problem defined by (1)–(2). It is known that $J_\epsilon = x_0'P_\epsilon x_0$ and that the optimal control which stabilizes the system and which achieves the optimal cost is given by $u = -(\epsilon^2 I + D'D)^{-1}(B'P_\epsilon + D'C)x$, where P_ϵ is the unique positive semi-definite solution to the following algebraic Riccati equation (ARE)

$$\begin{aligned} &[A - B(\epsilon^2 I + D'D)^{-1}D'C]'P_\epsilon \\ &+ P_\epsilon[A - B(\epsilon^2 I + D'D)^{-1}D'C] \\ &+ C'[I - D(\epsilon^2 I + D'D)^{-1}D']C \\ &- P_\epsilon B(\epsilon^2 I + D'D)^{-1}B'P_\epsilon = 0. \end{aligned} \tag{9}$$

It is easy to show (Kwakernaak and Sivan (1972b)) that P_ϵ monotonically decreases as ϵ goes to zero, and so the limit $P_0 := \lim_{\epsilon\to 0} P_\epsilon$ exists. The following result was proved in Scherzinger and Davison (1985); the same result for the case when $D = 0$ was obtained in Kwakernaak and Sivan (1972b) and Francis (1979).

Lemma 5. $P_0 = 0$ if and only if (A, B, C, D) is minimum phase and right-invertible.

For systems which do not satisfy the conditions given in Lemma 5, (Francis, 1979) characterized the null space of P_0, which is simply the set of all x_0 with $J_0 = 0$, and (Saberi and Sannuti, 1987) gave a complete decomposition of the state space in terms of the transient speed of the state trajectories. In the following, we will show that P_0 takes on a very simple form if (A, B, C, D) is a factorized realization of an arbitrary transfer matrix.

Lemma 6. Let (A, B, C, D) be a factorized realization of a transfer matrix. Then $P_\epsilon = \begin{bmatrix} I & 0 \\ 0 & P_{\epsilon 2} \end{bmatrix}$, where $P_{\epsilon 2}$ is the unique positive semi-definite solution to the ARE

$$\begin{aligned} &[A_2 - B_2(\epsilon^2 I + D_2'D_2)^{-1}D_2'C_2]'P_{\epsilon 2} \\ &+ P_{\epsilon 2}[A_2 - B_2(\epsilon^2 I + D_2'D_2)^{-1}D_2'C_2] \\ &+ C_2'[I - D_2(\epsilon^2 I + D_2'D_2)^{-1}D_2]C_2 \\ &- P_{\epsilon 2}B_2(\epsilon^2 I + D_2'D_2)^{-1}B_2P_{\epsilon 2} = 0. \end{aligned}$$

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be partitioned accordingly with the partition of A. Then the optimal control is given by $u = -(\epsilon^2 I + D_2' D_2)^{-1}(B_2' P_{\epsilon 2} + D_2' C_2)x_2$.

The proof of the first statement of this lemma is obtained simply by verifying that the given solution indeed satisfies ARE (9). The second statement follows by using Lemma 1. The details of the proof are dry algebra and are omitted.

A direct application of Lemma 6 leads to the following corollary.

Corollary 1. If (A, B, C, D) is a balanced realization of an inner transfer matrix, then $P_\epsilon = I$ and the optimal control is $u = 0$.

Since any minimal realization of a transfer matrix is similar to a balanced realization, this corollary implies that if (A, B, C, D) is a minimal realization of an inner matrix, then the optimal control of the system (1) under cost (2) is always zero, and thus the optimal performance is independent of ϵ. This is consistent with the well-known fact that cheap control asymptotically puts all the poles of the closed loop system to the mirror points of the system's zeros in $\mathbb{C}^+$ with respect to the imaginary axes. Since an inner matrix already has this property, no control is therefore needed to make it optimal.

Since (A_2, B_2, C_2, D_2) is a stabilizable and detectable realization of a minimum phase and right-invertible transfer matrix, the following theorem is obtained immediately from Lemmas 5 and 6. (In the statement of the theorem, we assume that P_ϵ is partitioned accordingly with the partition of A given in (8).)

Theorem 1. Let (A, B, C, D) be a factorized realization of a transfer matrix. Then

$$P_0 = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

4 FEEDFORWARD SERVOMECHANISM CONTROLLER

In the rest of this paper, we apply Theorem 1 to study various optimal servomechanism problems with transient performance measured by cheap quadratic functionals. Consider the servomechanism problem for system (3). Denote the zeros of (A, B, C, D) which are contained in $\mathbb{C}^+$ (if any) by $\lambda_1, \lambda_2, \ldots, \lambda_l$. To achieve clarity in the presentation, we assume in this and the next section that the disturbances are constant signals. The results will be extended to the sinusoidal signals in Section 6.

In order for the servomechanism problem to be solvable, it is necessary and sufficient to have the following assumption.

Assumption 1. Assume that $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ has full row rank.

Assumption 1 implies that system (3) is right-invertible and has no zero at the origin. Assume that disturbances ξ and η are measurable. Then under Assumption 1, the following feedforward controller solves the servomechanism problem:

$$u = Kx - \xi - [D - (C + DK)(A + BK)^{-1}B]^\dagger \eta, \tag{10}$$

where K is any matrix which makes $A + BK$ stable and $[\cdot]^\dagger = [\cdot]'([\cdot][\cdot]')^{-1}$ (Davison, 1973). The inverse involved exists due to Assumption 1.

Assume controller (10) is applied to the system (3). The closed loop stability implies that the input and the state of the system will approach constant values in the steady-state. Denote the steady-state values of the input and the state by $\bar{u}$ and $\bar{x}$, respectively. Then $\bar{u}$ and $\bar{x}$ must satisfy equations

$$\begin{aligned} 0 &= A\bar{x} + B(\bar{u} + \xi), \\ 0 &= C\bar{x} + D(\bar{u} + \xi) + \eta, \\ \bar{u} &= K\bar{x} - \xi \\ &\quad - [D - (C + DK)(A + BK)^{-1}B]^\dagger \eta. \end{aligned}$$

Let the transient part of the input and the state be denoted by $\tilde{u} := u - \bar{u}$ and $\tilde{x} := x - \bar{x}$, respectively. Then these values are governed by the following equations:

$$\begin{aligned} \dot{\tilde{x}} &= A\tilde{x} + B\tilde{u}, \quad \tilde{x}(0) = -\bar{x}, \\ y &= C\tilde{x} + D\tilde{u}, \\ \tilde{u} &= K\tilde{x}. \end{aligned}$$

These equations suggest that in order to have a good transient response, we can choose K to achieve the following optimal quadratic cost functional:

$$J_\epsilon = \min_{\tilde{u}} \int_0^t (y'y + \epsilon^2 \tilde{u}'\tilde{u})\, dt. \tag{11}$$

From the knowledge of the LQR problem, J_ϵ is a positive semi-definite quadratic form on $\tilde{x}(0) = -\bar{x}$, which in turn is linear in ξ and η. Hence, J_ϵ is a positive semi-definite quadratic form on ξ and η. Since J_ϵ monotonically decreases as ϵ^2 goes to zero, it follows that $J_0 := \lim_{\epsilon \to 0} J_\epsilon$ exists. The following theorem says that J_0 is a positive semi-definite quadratic form on η only, i.e. J_0 is independent of the input disturbance, and that a norm of J_0 can be simply

given by the locations of the zeros of system (3) in $\mathbb{C}^+$.

Theorem 2. $J_0 = \eta' H \eta$ for some positive semi-definite H and $\operatorname{tr} H = 2 \sum_{i=1}^{l} \frac{1}{\lambda_i}$.

Proof. The nature of the problem setup indicates that J_ϵ depends solely on the transfer matrix F of system (3). Therefore, we can assume that (A, B, C, D) is a factorized realization which is of the form of (8). Let $F_1(s) = D_1 + C_1(sI - A_1)^{-1}B_1$ and $F_2(s) = D_2 + C_2(sI - A_2)^{-1}B_2$. Since F is right-invertible by Assumption 1, it follows from Lemma 3 that F_1 must be square and that the poles of F_1 are $-\lambda_1, -\lambda_2, \ldots, -\lambda_l$.

It is known that $J_\epsilon = \bar{x}'(0)P_\epsilon \bar{x}(0) = \bar{x}'P_\epsilon \bar{x}$, where P_ϵ is the unique positive semi-definite solution of ARE (9). By Theorem 1, $P_0 := \lim_{\epsilon \to 0} P_\epsilon = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. Let $\bar{x}$ be partitioned as $\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$ according to the partition of A as given by (8); then $J_0 = \bar{x}_1'\bar{x}_1$.

Now assume that the closed loop system is at steady-state. The output of F_1 must be $-\eta$ and the output of F_2 is therefore $-F_1^{-1}(0)\eta$. It then follows that $\bar{x}_1 = A_1^{-1}B_1F_1^{-1}(0)\eta$. Let $H = F_1^{-1}(0)B_1'A_1'^{-1}A_1^{-1}B_1F_1^{-1}(0)$. Then $J_0 = \eta' H \eta$. Since the matrix $F_1(0)$ is unitary, it follows that

$$\operatorname{tr} H = \operatorname{tr}(B_1'A_1'^{-1}A_1^{-1}B_1) = \operatorname{tr}(A_1^{-1}B_1B_1'A_1'^{-1}).$$

By using Lemmas 1 and 3, we obtain

$$\begin{aligned} \operatorname{tr} H &= -\operatorname{tr}[A_1^{-1}(A_1 + A_1')A_1'^{-1}] \\ &= -2\operatorname{tr}(A_1^{-1}) = 2\sum_{i=1}^{l} \frac{1}{\lambda_i}. \qquad \square \end{aligned}$$

It is not a surprise that the feedforward controller (10) produces perfect control, i.e. $J_0 = 0$, for the case when only the input disturbance is present, even if the system (3) is non-minimum phase. In fact, the feedforward controller generates a signal which completely cancels the input disturbance. However for the case when the output disturbance is present, perfect control cannot be obtained for non-minimum phase systems, and a norm (or an averaging effect) of the optimal performance J_ϵ is now bounded from below by $2\sum_{i=1}^{l} \frac{1}{\lambda_i}$. This result shows that $2\sum_{i=1}^{l} \frac{1}{\lambda_i}$ can be considered as a quantitative measure of the degree of difficulty in solving the servomechanism problem for non-minimum phase systems with constant disturbances. This result also emphasizes the fact that not all non-minimum phase systems behave the same. A system with a small positive zero is more difficult to control than a system whose zeros in $\mathbb{C}^+$ are far away from the origin. On the other hand, a conjugate pair of complex zeros $\alpha \pm j\beta$ in $\mathbb{C}^+$ with $\alpha \ll |\beta|$ will not cause significant difficulty in control since its contribution to the limiting performance J_0 is $\frac{4\alpha}{\alpha^2 + \beta^2}$. This later phenomenon has been observed in Davison and Gesing (1985) in the control design for a large flexible space structure. A similar result to Theorem 2 was given in Morari and Zafiriou (1989) for SISO systems by using frequency domain techniques.

5 ROBUST SERVOMECHANISM CONTROLLER

In practical applications, the parameters of system (3) are always somewhat uncertain and we often do not have access to the disturbances. In either case, controller (10) cannot generally be used. To overcome these difficulties, a robust servomechanism controller has been proposed to solve the servomechanism problem (Davison, 1976). The robust servomechanism controller does not require the disturbances to be measured, but must contain a servocompensator which contains the modes of the disturbances. A significant advantage of the robust servomechanism controller over the feedforward controller is that tracking and disturbance rejection occur for all perturbations in the system provided only that the perturbed closed loop system remains stable.

Assume again that Assumption 1 holds. The robust servomechanism controller for system (3) with constant disturbances can take the following form

$$\begin{aligned} \dot{z} &= y, \quad z(0) = 0, \\ u &= K_0 x + K z, \end{aligned} \tag{12}$$

where $[K_0 \quad K]$ is chosen to stabilize matrix $\begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} + \begin{bmatrix} B \\ D \end{bmatrix} [K_0 \quad K]$.

On combining the original system and the servo-compensator (the integrator), the augmented system with input u and output z is then described by the state-space equation:

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} B \\ D \end{bmatrix} (u + \xi) + \begin{bmatrix} 0 \\ I \end{bmatrix} \eta,$$

$$z = [0 \quad I] \begin{bmatrix} x \\ z \end{bmatrix}, \quad \begin{bmatrix} x(0) \\ z(0) \end{bmatrix} =$$

The action of the controller becomes that of a state feedback, i.e. $u = [K_0 \quad K]\begin{bmatrix} x \\ z \end{bmatrix}$. Define new variables: $\tilde{x} := \dot{x}$, $\tilde{z} := \dot{z}$, $\tilde{u} := \dot{u}$. On noticing that $\dot{z} = y$, the augmented system then becomes

$$\begin{bmatrix} \dot{\tilde{x}} \\ \dot{\tilde{z}} \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{z} \end{bmatrix} + \begin{bmatrix} B \\ D \end{bmatrix}\tilde{u},$$

$$\begin{bmatrix} \tilde{x}(0) \\ \tilde{z}(0) \end{bmatrix} = \begin{bmatrix} B & 0 \\ D & I \end{bmatrix}\begin{bmatrix} \xi \\ \eta \end{bmatrix}, \tag{13}$$

$$y = [0 \quad I]\begin{bmatrix} \tilde{x} \\ \tilde{z} \end{bmatrix},$$

and the controller becomes $\tilde{u} = [K_0 \quad K]\begin{bmatrix} \tilde{x} \\ \tilde{z} \end{bmatrix}$. It can be easily shown that system

$$\left(\begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix}, \begin{bmatrix} B \\ D \end{bmatrix}, [0 \quad I], 0\right),$$

is always stabilizable and detectable under Assumption 1 and the assumption that (A, B, C, D) is stabilizable and detectable.

This suggests that in order to have a good transient response, we can choose $[K_0 \quad K]$ to achieve the following optimal quadratic cost functional

$$J_\epsilon = \min_u \int_0^\infty (y'y + \epsilon^2\tilde{u}'\tilde{u})\,dt. \tag{14}$$

By the same argument as made in the last section, we see that J_ϵ is a positive semi-definite quadratic form on ξ and η, and that $J_0 := \lim_{\epsilon\to 0} J_\epsilon$ exists. The statement of the following theorem is exactly the same as in Theorem 2.

Theorem 3. $J_0 = \eta' H \eta$ for some real positive semi-definite H and $\operatorname{tr} H = 2\sum_{i=1}^{l} \frac{1}{\lambda_i}$.

Proof. Again J_ϵ depends solely on the transfer matrix F of system (3). Therefore (A, B, C, D) can be assumed to be a factorized realization of the form of (8). Consequently (13) can be rewritten as

$$\begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{x}}_2 \\ \dot{\tilde{z}} \end{bmatrix} = \begin{bmatrix} A_1 & B_1C_2 & 0 \\ 0 & A_2 & 0 \\ C_1 & D_1C_2 & 0 \end{bmatrix}\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{z} \end{bmatrix} + \begin{bmatrix} B_1D_2 \\ B_2 \\ D_1D_2 \end{bmatrix}u,$$

$$y = [0 \quad 0 \quad I]\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{z} \end{bmatrix}, \tag{15}$$

with initial condition

$$\begin{bmatrix} \tilde{x}_1(0) \\ \tilde{x}_2(0) \\ \tilde{z}(0) \end{bmatrix} = \begin{bmatrix} B_1D_2 & 0 \\ B_2 & 0 \\ D_1D_2 & I \end{bmatrix}\begin{bmatrix} \xi \\ \eta \end{bmatrix}.$$

Let

$$T = \begin{bmatrix} A_1 & B_1 \\ 0 & 0 \\ C_1 & D_1 \end{bmatrix}$$

and

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{z} \end{bmatrix} = T \begin{bmatrix} \hat{x}_1 \\ \hat{z} \\ \hat{x}_2 \end{bmatrix}.$$

Assumption 1 guarantees that T is invertible. System (15) can then be transformed to

$$\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{z}} \\ \dot{\hat{x}}_2 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 & 0 \\ 0 & 0 & C_2 \\ 0 & 0 & A_2 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{z} \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ D_2 \\ B_2 \end{bmatrix}\tilde{u},$$

$$y = [C_1 \quad D_1 \quad 0]\begin{bmatrix} \hat{x}_1 \\ \hat{z} \\ \hat{x}_2 \end{bmatrix}, \tag{16}$$

and the initial condition then becomes

$$\begin{bmatrix} \hat{x}_1(0) \\ \hat{z}(0) \\ \hat{x}_2(0) \end{bmatrix} \quad \begin{bmatrix} B_1D_2 & 0 \\ B_2 & 0 \\ D_1D_2 & I \end{bmatrix}\begin{bmatrix} \xi \\ \eta \end{bmatrix}.$$

Note that (16) is a factorized realization with (A_1, B_1, C_1, D_1) being inner and $\left(\begin{bmatrix} 0 & C_2 \\ 0 & A_2 \end{bmatrix}, \begin{bmatrix} D_2 \\ B_2 \end{bmatrix}, [I \quad 0], 0\right)$ being minimum phase and right-invertible. Let P_ϵ be the unique positive semi-definite solution to the ARE associated with system (16) and optimal cost functional (14). Theorem 1 leads to

$$\lim_{\epsilon\to 0} P_\epsilon = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and $J_0 = \hat{x}_1'(0)\hat{x}_1(0)$. Direct calculation shows that

$$\hat{x}_1(0) = [I \quad 0 \quad 0]T^{-1}\begin{bmatrix} B_1D_2 & 0 \\ B_2 & 0 \\ D_1D_2 & I \end{bmatrix}\begin{bmatrix} \xi \\ \eta \end{bmatrix}$$

$$= -A_1^{-1}B_1F_1^{-1}(0)\eta.$$

where $F_1(0) = D_1 - C_1A_1^{-1}B_1$. Let $H = F_1'^{-1}(0)B_1'A_1'^{-1}A_1^{-1}B_1F_1^{-1}(0)$. Then $J_0 = \eta' H \eta$.

The rest of the proof now proceeds in exactly the same way as done in the last part of the proof of Theorem 2. □

Two interesting points can be observed on comparing Theorem 3 with Theorem 2. Firstly, although we can no longer completely cancel the input disturbance when the robust servomechanism controller is used (as was done for

the feedforward controller) perfect control still occurs for the case when only the input disturbance is present even if the system is non-minimum phase. In other words, no matter whether the system is minimum phase or not, the robust servomechanism controller's reaction to the input disturbance can be made arbitrarily fast. This result is perhaps somewhat surprising. Secondly, we are concerned if the use of the robust servomechanism controller sacrifices the potential performance of the controlled system, in comparison to the use of the feedforward controller. Since the variable $\bar{u}$ in (11) and (14) have different physical meanings, it may appear that the norms of J_0 given in Theorem 2 and Theorem 3 are incomparable. However, it has been shown in Kwakernaak and Sivan (1972b) that if we denote the outputs of the optimal system in both cases by y_ϵ, then $J_0 = \lim_{\epsilon \to 0} \int_0^\infty y_\epsilon' y_\epsilon \, dt$, i.e. when ϵ is small, J_ϵ essentially contains only the output term. Therefore, in both cases, the limiting transient speed of the output, measured by a norm of J_0, are the same. In other words, the use of the robust servomechanism controller will not lead to a significant loss in the potential limiting performance of the system.

6 SINUSOIDAL DISTURBANCES

In this section, we extend the results obtained in the last two sections to the case when the disturbances are sinusoidal signals. We will obtain results which are in the same spirit as in Theorems 2–3.

Assume that the disturbances in system (3) are now of the following form

$$\xi(t) = \xi_{e1} \sin \omega t + \xi_{e2} \cos \omega t, \qquad (17)$$

$$\eta(t) = \eta_{e1} \sin \omega t + \eta_{e2} \cos \omega t, \qquad (18)$$

where $\xi_e := \begin{bmatrix} \xi_{e1} \\ \xi_{e2} \end{bmatrix}$ and $\eta_e := \begin{bmatrix} \eta_{e1} \\ \eta_{e2} \end{bmatrix}$ are real constant vectors. In order for the servomechanism problem to be solvable for system (3) with disturbances of the form (17)–(18), it is necessary and sufficient to have the following assumption.

Assumption 2. Assume that $\begin{bmatrix} A - j\omega I & B \\ C & D \end{bmatrix}$ has full row rank.

Assumption 2 implies that system (3) is right-invertible and has no zero at $j\omega$.

First, we consider the ideal case when the system parameters (A, B, C, D) are exactly known, and the disturbances as well as their derivatives are available for measurement. In this case, the following feedforward controller solves the servomechanism problem:

$$\begin{aligned} u = Kx - \xi - \mathrm{Re}\, \{&[D + (C + DK) \\ &\times (j\omega I - A - BK)^{-1} B]^\dagger\} \eta \\ &- \frac{1}{\omega} \mathrm{Im}\, \{[D + (C + DK) \\ &\times (j\omega I - A - BK)^{-1} B]^\dagger\} \dot{\eta}, \end{aligned} \qquad (19)$$

where K is any matrix which makes $A + BK$ stable and $[\cdot]^\dagger = [\cdot]'([\cdot][\cdot]')^{-1}$ (Davison, 1973). The inverse involved exists due to Assumption 2.

Assume controller (19) is applied to system (3), and assume that the steady-state signals of the input and the state are given by $\bar{u}$ and $\bar{x}$, respectively. Then $\bar{u}$ and $\bar{x}$ must satisfy equations:

$$\begin{aligned} \dot{\bar{x}} &= A\bar{x} + B\bar{u} + B\xi, \\ 0 &= C\bar{x} + D\bar{u} + D\xi + \eta, \\ \bar{u} &= K\bar{x} - \xi - \mathrm{Re}\, \{[D + (C + DK) \\ &\quad \times (j\omega I - A - BK)^{-1} B]^\dagger\} \eta \\ &\quad - \frac{1}{\omega} \mathrm{Im}\, \{[D + (C + DK) \\ &\quad \times (j\omega I - A - BK)^{-1} B]^\dagger\} \dot{\eta}. \end{aligned}$$

Let the transient part of the input and the state be denoted by $\tilde{u} := u - \bar{u}$ and $\tilde{x} := x - \bar{x}$, respectively. Then these values are governed by the following equations:

$$\begin{aligned} \dot{\tilde{x}} &= A\tilde{x} + B\tilde{u}, \quad \tilde{x}(0) = -\bar{x}(0), \\ y &= C\tilde{x} + D\tilde{u}, \\ \tilde{u} &= K\tilde{x}. \end{aligned}$$

These equations suggest that in order to have a good transient response, we can choose K to achieve the following optimal quadratic cost functional

$$J_\epsilon = \min_u \int_0^\infty (y'y + \epsilon^2 \tilde{u}'\tilde{u}) \, dt. \qquad (20)$$

From the knowledge of the LQR problem, J_ϵ is a positive semi-definite quadratic form on $\tilde{x}(0) = -\bar{x}(0)$, which in turn depends linearly on ξ_e and η_e. Hence J_ϵ is a positive semidefinite quadratic form on ξ_e and η_e. Since J_ϵ decreases monotonically as ϵ^2 goes to zero, it follows that $J_0 := \lim_{\epsilon \to 0} J_\epsilon$ exists. The following theorem says that J_0 is a positive semidefinite quadratic form on η_e only and a norm of J_0 is given by a simple expression involving only the zeros of the system (3) in $\mathbb{C}^+$.

Theorem 4. $J_0 = \eta_e' M \eta_e$ for some positive semi-definite M and

$$\operatorname{tr} M = \sum_{i=1}^{l} \left(\frac{1}{\lambda_i - j\omega} + \frac{1}{\lambda_i + j\omega} \right).$$

Proof. Similar to the constant disturbance case, we can assume that (A, B, C, D) is a factorized realization of the form of (8). Let $F_1(s) = D_1 + C_1(sI - A_1)^{-1}B_1$ and $F_2(s) = D_2 + C_2(sI - A_2)^{-1}B_2$. Since system (3) is assumed to be right-invertible, it follows from Lemma 3 that F_1 must be square and that the poles of F_1 are given by $-\lambda_1, -\lambda_2, \ldots, -\lambda_l$.

It is known that $J_\epsilon = \bar{x}'(0)P_\epsilon \bar{x}(0) = \bar{x}'(0)P_\epsilon \bar{x}(0)$ where P_ϵ is the unique positive semidefinite solution of ARE (9). By Theorem 1, $P_0 := \lim_{\epsilon \to 0} P_\epsilon = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. Let $\bar{x}$ be partitioned as $\bar{x} =$

according to the partition of A given by (8); then $J_0 = \bar{x}_1'(0)\bar{x}_1(0)$.

Now assume that the closed loop system is at steady-state. The output of F_1 must be $-\eta$. To obtain the state $\bar{x}_1$ of F_1, we use $\mathscr{L}(\cdot)$ to denote the Laplace transform operator. Then $\bar{x}_1$ is related to η by

$$\mathscr{L}(\bar{x}_1) = -(sI - A_1)^{-1}B_1F_1^{-1}(s)\mathscr{L}(\eta).$$

Let $L(s) = (sI - A_1)^{-1}B_1F_1^{-1}(s)$. Steady state sinusoidal analysis tells us that

$$\bar{x}_1 = -\operatorname{Re} L(j\omega)\eta - \frac{1}{\omega} \operatorname{Im} L(j\omega)\dot{\eta}.$$

Hence,

$$\bar{x}_1(0) = -\operatorname{Re} L(j\omega)\eta_{e2} - \operatorname{Im} L(j\omega)\eta_{e1}.$$

Let

$$M = \begin{bmatrix} \operatorname{Im} L(j\omega) \\ \operatorname{Re} L(j\omega) \end{bmatrix}' [\operatorname{Im} L(j\omega) \ \operatorname{Re} L(j\omega)].$$

Then $J_0 = \eta_e' M \eta_e$ and

$$\begin{aligned}
\operatorname{tr} M &= \operatorname{tr}[L^*(j\omega)L(j\omega)] \\
&= \operatorname{tr}[F_1^{*-1}(j\omega)B_1'(j\omega I - A_1)^{*-1} \\
&\quad \times (j\omega I - A_1)^{-1}B_1F_1^{-1}(j\omega)] \\
&= \operatorname{tr}[B_1'(j\omega I - A_1)^{*-1}(j\omega I - A_1)^{-1}B_1] \\
&= \operatorname{tr}[(j\omega I - A_1)^{-1}B_1B_1'(-j\omega I - A_1')^{-1}].
\end{aligned}$$

By using Lemmas 1 and 3, we obtain

$$\begin{aligned}
\operatorname{tr} M &= \operatorname{tr}[(j\omega I - A_1)^{-1} \\
&\quad \times (j\omega I - A_1 - j\omega I - A_1')(-j\omega I - A_1')^{-1}] \\
&= \operatorname{tr}[(-j\omega I - A_1')^{-1} + (j\omega I - A_1)^{-1}] \\
&= \sum_{i=1}^{l} \left(\frac{1}{\lambda_i - j\omega} + \frac{1}{\lambda_i - j\omega} \right).
\end{aligned}$$ □

In this case, we see that a system with zeros in $\mathbb{C}^+$ close to $j\omega$, where ω is the frequency of the disturbances, is more difficult to control than a system whose zeros in $\mathbb{C}^+$ are far away from $j\omega$.

We can see, as in the constant disturbance case, that the feedforward controller (19) cannot be used either when the system parameters (A, B, C, D) are uncertain or when the disturbances are not available for measurement. The robust servomechanism controller does not have these disadvantages. The robust servomechanism controller for disturbance signals of the form (17)–(18) can take the following form (Davison, 1976)

$$\begin{aligned}
\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} &= \begin{bmatrix} 0 & -\omega^2 I \\ I & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} y, \\
& z_1(0) \\
& z_2(0), \\
u &= K_0 x + K_1 z_1 + K_2 z_2,
\end{aligned} \tag{21}$$

where $[K_0 \ K_1 \ K_2]$ is chosen to stabilize the following matrix

$$\begin{matrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{matrix} + \begin{matrix} B \\ D \\ 0 \end{matrix} [K_0 \quad K_1 \quad K_2].$$

The augmented system (with input u and output z_2) is then described by the state-space equation

$$\begin{bmatrix} \dot{x} \\ \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} x \\ z_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} B \\ D \\ 0 \end{bmatrix} (u + \xi) + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \eta, \quad \begin{bmatrix} x(0) \\ z_1(0) \\ z_2(0) \end{bmatrix} = 0,$$

$$z_2 = [0 \quad 0 \quad I] \begin{bmatrix} x \\ z_1 \\ z_2 \end{bmatrix}.$$

Define new variables $\bar{x} := \ddot{x} + \omega^2 x$, $\bar{z}_1 := \ddot{z}_1 + \omega^2 z_1$, $\bar{z}_2 := \ddot{z}_2 + \omega^2 z_2$, and $\bar{u} := \ddot{u} + \omega^2 u$. On noting that $\ddot{z}_2 + \omega^2 z_2 = y$, the augmented system then becomes

$$\begin{aligned}
\begin{bmatrix} \dot{\bar{x}} \\ \dot{\bar{z}}_1 \\ \dot{\bar{z}}_2 \end{bmatrix} &= \begin{bmatrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{z}_1 \\ \bar{z}_2 \end{bmatrix} + \begin{bmatrix} B \\ D \\ 0 \end{bmatrix} \\
y &= [0 \quad 0 \quad I] \begin{bmatrix} \bar{x} \\ \bar{z}_1 \\ \bar{z}_2 \end{bmatrix},
\end{aligned} \tag{22}$$

where $\left(\begin{bmatrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{bmatrix}, \begin{bmatrix} B \\ D \\ 0 \end{bmatrix}, [0 \quad 0 \quad I], 0 \right.$

is always stabilizable and detectable under Assumption 2 and the assumption that (A, B, C, D) is stabilizable and detectable.

The initial condition of system (22) is given by

$$\begin{bmatrix} \bar{x}(0) \\ \bar{z}_1(0) \\ \bar{z}_2(0) \end{bmatrix} = \begin{bmatrix} \ddot{x}(0) \\ \ddot{z}_1(0) \\ \ddot{z}_2(0) \end{bmatrix} = \begin{bmatrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} \dot{x}(0) \\ \dot{z}_1(0) \\ \dot{z}_2(0) \end{bmatrix} + \begin{bmatrix} B \\ D \\ 0 \end{bmatrix} [\dot{u}(0) + \dot{\xi}(0)] + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \dot{\eta}(0)$$

$$= \begin{bmatrix} A & 0 & 0 \\ C & 0 & -\omega^2 I \\ 0 & I & 0 \end{bmatrix} \times \left\{ \begin{bmatrix} B \\ D \\ 0 \end{bmatrix} [u(0) + \xi(0)] + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \eta(0) \right\} + \begin{bmatrix} B \\ D \\ 0 \end{bmatrix} [\dot{u}(0) + \dot{\xi}(0)] + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \dot{\eta}(0).$$

These equations suggest that in order to have a good transient response, we can choose $[K_0 \ K_1 \ K_2]$ to achieve the following optimal quadratic cost functional

$$J_\epsilon = \min_u \int_0^\infty (y'y + \epsilon^2 \bar{u}'\bar{u})\, dt. \tag{23}$$

Again the optimal cost J_ϵ is a positive semi-definite quadratic form on ξ_e and η_e, and $J_0 = \lim J_\epsilon$ exists. The statement of the next theorem is exactly the same as that of Theorem 4.

Theorem 5. $J_0 = \eta_e' M \eta_e$ for some positive semi-definite M and

$$\operatorname{tr} M = \sum_{i=1}^{l} \left(\frac{1}{\lambda_i - j\omega} + \frac{1}{\lambda_i + j\omega} \right).$$

Proof. Again we can assume that (A, B, C, D) factorized realization of P which is of the of (8). Consequently (22) can be written as

$$\begin{bmatrix} \dot{\bar{x}}_1 \\ \dot{\bar{x}}_2 \\ \dot{\bar{z}}_1 \\ \dot{\bar{z}}_2 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 C_2 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ C_1 & D_1 C_2 & 0 & -\omega^2 I \\ 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{z}_1 \\ \bar{z}_2 \end{bmatrix} + \begin{bmatrix} B_1 D_2 \\ B_2 \\ D_1 D_2 \\ 0 \end{bmatrix} \bar{u}, \tag{24}$$

$$y = [0 \ \ 0 \ \ 0 \ \ I] \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{z}_1 \\ \bar{z}_2 \end{bmatrix}$$

Let

$$T = \begin{bmatrix} A_1^2 + \omega^2 I & A_1 B_1 & B_1 & 0 \\ 0 & 0 & 0 & I \\ C_1 A_1 & C_1 B_1 & D_1 & 0 \\ C_1 & D_1 & 0 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{z}_1 \\ \bar{z}_2 \end{bmatrix} = T \begin{bmatrix} \hat{x}_1 \\ \hat{z}_1 \\ \hat{z}_2 \\ \hat{x}_2 \end{bmatrix}.$$

Assumption 2 guarantees that T is invertible. System (24) is then transformed to

$$\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{z}}_1 \\ \dot{\hat{z}}_2 \\ \dot{\hat{x}}_2 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & -\omega^2 I & 0 & C_2 \\ 0 & 0 & 0 & A_2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{z}_1 \\ \hat{z}_2 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ D_2 \\ B_2 \end{bmatrix} \bar{u}, \tag{25}$$

$$y = [C_1 \ \ D_1 \ \ 0 \ \ 0] \begin{bmatrix} \hat{x}_1 \\ \hat{z}_1 \\ \hat{z}_2 \\ \hat{x}_2 \end{bmatrix}$$

System (25) is in the factorized form with (A_1, B_1, C_1, D_1) being inner and

$$\left(\begin{bmatrix} 0 & I & 0 \\ -\omega^2 I & 0 & C_2 \\ 0 & 0 & A_2 \end{bmatrix}, \begin{bmatrix} 0 \\ D_2 \\ B_2 \end{bmatrix}, [I \ \ 0 \ \ 0], 0 \right)$$

being minimum phase and right-invertible. Let P_ϵ be the unique positive semi-definite solution of the ARE associated with system (25) and the optimal cost functional (23). Theorem 1 leads to

$$P_0 := \lim_{\epsilon \to 0} P_\epsilon = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and $J_0 = \hat{x}_1'(0)\hat{x}_1(0)$.

The initial condition $\hat{x}_1(0)$ can be obtained as:

$$\hat{x}_1(0) = [I \quad 0 \quad 0 \quad 0]\begin{bmatrix} \hat{x}_1(0) \\ \hat{z}_1(0) \\ \hat{z}_2(0) \\ \hat{x}_2(0) \end{bmatrix}$$

$$= [I \quad 0 \quad 0 \quad 0]T^{-1}$$

$$\times\left\{\begin{bmatrix} A_1 & B_1C_2 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ C_1 & D_1C_2 & 0 & -\omega^2 I \\ 0 & 0 & I & 0 \end{bmatrix}\right.$$

$$\times\left(\begin{bmatrix} B_1D_2 \\ B_2 \\ D_1D_2 \\ 0 \end{bmatrix}[u(0)+\xi(0)] + \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix}\eta(0)\right)$$

$$\left. + \begin{bmatrix} B_1D_2 \\ B_2 \\ D_1D_2 \\ 0 \end{bmatrix}[\dot{u}(0)+\dot{\xi}(0)] + \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix}\dot{\eta}(0)\right\}.$$

Straightforward computation leads to:

$$\hat{x}_1(0) = -[(A_1 - B_1D_1^{-1}C_1)^2 + \omega^2 I]^{-1}$$
$$\times [(A_1 - B_1D_1^{-1}C_1)B_1D_1^{-1}\eta(0)$$
$$+ B_1D_1^{-1}\dot{\eta}(0)]$$
$$= -[(A_1 - B_1D_1^{-1}C_1)^2 + \omega^2 I]^{-1}$$
$$\times [(A_1 - B_1D_1^{-1}C_1)B_1D_1^{-1}\eta_{c2}$$
$$+ \omega B_1D_1^{-1}\eta_{c1}].$$

Let $L(s) = (sI - A_1 + B_1D_1^{-1}C_1)^{-1}B_1D_1^{-1}$. Then simple algebra shows that

$$L(s) = (sI - A_1)^{-1}B_1$$
$$\times [D_1 + C_1(sI - A_1)^{-1}B_1]^{-1}$$
$$= (sI - A_1)^{-1}B_1F_1^{-1}(s),$$

and

$$\hat{x}_1(0) = -\mathrm{Re}\,[L(j\omega)]\eta_{c2} - \mathrm{Im}\,[L(j\omega)]\eta_{c1}.$$

Let

$$M = \begin{bmatrix} \mathrm{Im}\,L(j\omega) \\ \mathrm{Re}\,L(j\omega) \end{bmatrix}' [\mathrm{Im}\,L(j\omega)\ \mathrm{Re}\,L(j\omega)].$$

Then $J_0 = \eta_c' M \eta_c$. The rest of this proof is the same as the last part of the proof of Theorem 4. □

Again we have two similar observations as in the constant disturbance case. Firstly, perfect control still occurs for the case when only the input disturbance is present, even if the system is non-minimum phase. Secondly, the use of the robust servomechanism controller does not lead to a significant loss in the potential limiting performance of the system, compared to the feedforward controller case.

7. AN EXAMPLE

An experimental flexible beam system is described by the following transfer function (MacLean, 1990)

$$F(s) = \frac{8.26s^4 - 1.66s^3 - 2878s^2 + 453s + 95400}{5s^6 + 4.83s^5 + 2312s^4 + 488s^3 + 60657s^2 + 40.5s},$$

which has a state space model of the form (3) given by

$$A = \begin{bmatrix} -0.996 & -463 & -97.8 & -12131 & -8.11 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$C = [0 \quad 1.65 \quad -0.331 \quad -576 \quad 90.6 \quad 19080],$$
$$D = 0.$$

Note that this is not in the factorized form as in (8).

This system has two zeros in $\mathbb{C}^+$: $\lambda_1 = 6.18$ and $\lambda_2 = 17.7$. Our purpose is to solve the servomechanism problem for this system with respect to a constant reference signal y_{ref} (assuming that there are no external disturbances present). As we have discussed in Section 1, we can consider $y - y_{\mathrm{ref}}$ as the output and $-y_{\mathrm{ref}}$ as the output disturbance for this problem.

Let us first apply the feedforward controller of the form (10) (with $D = 0$):

$$u = K_\epsilon x + [-C(A + BK_\epsilon)^{-1}B]^{-1}y_{\mathrm{ref}},$$

where K_ϵ is chosen to achieve the optimal cost

$$J_\epsilon = \min_u \int_0^\infty [(y - y_{\mathrm{ref}})'(y - y_{\mathrm{ref}}) + \epsilon^2(u - \bar{u})'(u - \bar{u})]\,dt.$$

Assume that the system has a zero initial condition. Let $\bar{x}_\epsilon$ be the steady state value of the state variable; then $\bar{x}_\epsilon = -(A + BK_\epsilon)^{-1}B[-$

TABLE 1. COMPARISON OF OPTIMAL COSTS FOR FEEDFORWARD CONTROLLER CASE

ϵ	J_ϵ	$J_{y\epsilon}$	$J_{u\epsilon}$
1	$1.21y^2_{\rm ref}$	$0.957y^2_{\rm ref}$	$2.54\times10^{-1}y^2_{\rm ref}$
10^{-1}	$0.640y^2_{\rm ref}$	$0.592y^2_{\rm ref}$	$4.86y^2_{\rm ref}$
10^{-2}	$0.513y^2_{\rm ref}$	$0.499y^2_{\rm ref}$	$1.39\times10^{2}y^2_{\rm ref}$
10^{-3}	$0.468y^2_{\rm ref}$	$0.462y^2_{\rm ref}$	$6.70\times10^{3}y^2_{\rm ref}$
10^{-4}	$0.448y^2_{\rm ref}$	$0.445y^2_{\rm ref}$	$2.62\times10^{5}y^2_{\rm ref}$

$C(A+BK_\epsilon)^{-1}B]^{-1}y_{\rm ref}$ and $J_\epsilon=(-\bar{x}_\epsilon)'P_\epsilon(-\bar{x}_\epsilon)$, where P_ϵ is the unique positive semi-definite solution to ARE (9) (with $D=0$). From Theorem 2, we obtain $J_\epsilon\to2\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)y^2_{\rm ref}=0.437y^2_{\rm ref}$ as $\epsilon\to0$. Table 1 gives the computed value of J_ϵ for several different values of ϵ. For the convenience of comparison, Table 1 also gives

$$J_{y\epsilon}:=\int_0^\infty(y_\epsilon-y_{\rm ref})'(y_\epsilon-y_{\rm ref})\,{\rm d}t,$$

and

$$J_{u\epsilon}:=\int_0^\infty(u_\epsilon-\bar{u})'(u_\epsilon-\bar{u})\,{\rm d}t,$$

where y_ϵ and u_ϵ are the output and control trajectories of the optimally controlled system. The quantities $J_{y\epsilon}$ and $J_{u\epsilon}$ can be obtained as ${\rm tr}\,(CL_\epsilon C')$ and ${\rm tr}\,(K_\epsilon L_\epsilon K_\epsilon')$, respectively, where L_ϵ is the solution of the following Lyapunov equation

$$(A+BK_\epsilon)L_\epsilon+L_\epsilon(A+BK_\epsilon)'+\bar{x}_\epsilon\bar{x}_\epsilon'=0.$$

It is seen that as $\epsilon\to0$, the computed value of J_ϵ is approaching the limiting cost $J_0=0.437y^2_{\rm ref}$ obtained from Theorem 2.

Now let us apply the robust controller

$$\dot{z}=y-y_{\rm ref},$$
$$u=K_{0\epsilon}x+K_\epsilon z,$$

where $[K_{0\epsilon}\ \ K_\epsilon]$ is chosen to achieve the optimal cost

$$J_\epsilon=\min_u\int_0^\infty[(y-y_{\rm ref})'(y-y_{\rm ref})+\epsilon^2\dot{u}'\dot{u}]\,{\rm d}t.$$

We know from Section 5 that $J_\epsilon=[0\quad -y_{\rm ref}]P_\epsilon\begin{bmatrix}0\\-y_{\rm ref}\end{bmatrix}$ where P_ϵ is the unique positive semi-definite solution of the following ARE:

$$\begin{bmatrix}A'&C'\\0&0\end{bmatrix}P_\epsilon+P_\epsilon\begin{bmatrix}A&0\\C&0\end{bmatrix}+\begin{bmatrix}0\\I\end{bmatrix}[0\quad I]-\frac{1}{\epsilon^2}P_\epsilon\begin{bmatrix}B\\0\end{bmatrix}[B'\quad0]P_\epsilon=0.$$

TABLE 2. COMPARISON OF OPTIMAL COSTS FOR ROBUST CONTROLLER CASE

	J_ϵ	$J_{y\epsilon}$	$J_{u\epsilon}$
1	$1.78y^2_{\rm ref}$	$1.50y^2_{\rm ref}$	$2.76\times10^{-1}y^2_{\rm ref}$
10^{-1}	$0.962y^2_{\rm ref}$	$0.867y^2_{\rm ref}$	$9.49y^2_{\rm ref}$
10^{-2}	$0.690y^2_{\rm ref}$	$0.650y^2_{\rm ref}$	$4.00\times10^{2}y^2_{\rm ref}$
10^{-3}	$0.563y^2_{\rm ref}$	$0.547y^2_{\rm ref}$	$1.63\times10^{4}y^2_{\rm ref}$
10^{-4}	$0.506y^2_{\rm ref}$	$0.497y^2_{\rm ref}$	$9.59\times10^{5}y^2_{\rm ref}$
10^{-5}	$0.472y^2_{\rm ref}$	$0.447y^2_{\rm ref}$	$5.48\times10^{7}y^2_{\rm ref}$

From Theorem 3, we again obtain that $J_\epsilon\to0.437y^2_{\rm ref}$ as $\epsilon\to0$. Table 2 gives the computed value of J_ϵ for several different values of ϵ. Table 2 also gives

$$J_{y\epsilon}:=\int_0^\infty(y_\epsilon-y_{\rm ref})'(y_\epsilon-y_{\rm ref})\,{\rm d}t,$$

and

$$J_{u\epsilon}:=\int_0^\infty\dot{u}_\epsilon'\dot{u}_\epsilon\,{\rm d}t,$$

where y_ϵ and u_ϵ are the output and control trajectories of the optimally controlled system. The quantities $J_{y\epsilon}$ and $J_{u\epsilon}$ can be obtained as ${\rm tr}\,([0\ \ I]L_\epsilon[0\ \ I]')$ and ${\rm tr}\,([K_{0\epsilon}\ \ K_\epsilon]L_\epsilon[K_{0\epsilon}\ \ K_\epsilon]')$, respectively, where L_ϵ is the solution of the following Lyapunov equation

$$\left(\begin{bmatrix}A&0\\C&0\end{bmatrix}+\begin{bmatrix}B\\0\end{bmatrix}[K_{0\epsilon}\quad K_\epsilon]\right)L$$
$$+L_\epsilon\left(\begin{bmatrix}A&0\\C&0\end{bmatrix}+\begin{bmatrix}B\\0\end{bmatrix}[K_{0\epsilon}\quad K_\epsilon]\right)'$$

$$[0$$

Again, it is seen that as $\epsilon\to0$, the computed value of J_ϵ is approaching the limiting cost $J_0=0.437y^2_{\rm ref}$ obtained from Theorem 3.

8. CONCLUSION

This paper considers the cheap regulator problem and the cheap optimal servomechanism problems for systems which may be non-minimum phase. The basic tool used is a factorization which factorizes an arbitrary system into the product of an inner system and a right-invertible minimum phase system. Based on this factorization, the study of an arbitrary system can be decomposed into the study of an inner system and the study of a right-invertible minimum phase system. The cheap control problem of an inner system becomes easy to analyse by exploiting various properties of inner matrices, while the cheap control problem of a right-invertible minimum phase system has already been intensively studied.

A novel contribution of this paper is the establishment of the fact that the number and the locations of the zeros of a system in the open right half of the complex plane, are crucial factors which determine the best attainable closed loop performance of the system. In particular, it is shown that the fundamental design limitations on the closed loop performance of the servomechanism problem can be completely characterized by the number and the locations of the zeros of the open loop system which lie in the open right half of the complex plane. This design limitation can be used to evaluate an open loop system, i.e. to determine whether the system is "inherently hard to control", and to assess a given closed loop design, i.e. to determine how near the closed loop system's performance is from the best attainable.

REFERENCES

Cheng, V. H. L. and C. A. Desoer (1980). Limitations on the closed loop transfer function due to right half plane transmission zeros of the plant. *IEEE Trans. Aut. Control*, **AC-25**, 1218–1220.

Davison, E. J. (1973). The feedforward control of linear multivariable time invariant systems. *Automatica*, **9**, 561–573.

Davison, E. J. (1976). The robust control of a servomechanism problem for linear time invariant multivariable systems. *IEEE Trans. Aut. Control*, **AC-21**, 25–34.

Davison, E. J. and S. G. Chow (1977). Perfect control in linear time invariant multivariable systems: the control inequality principle. In E. Fallside (Ed.), *Control System Design by Pole Zero Assignment*. Academic Press, London.

Davison, E. J. and W. Gesing (1985). Robust attitude and shape control of third generation spacecraft. *Advances in the Astronautical Sciences*, **58**, 431–450.

Davison, E. J. and B. M. Scherzinger (1987). Perfect control of the robust servomechanism problem. *IEEE Trans. Aut. Control*, **AC-32**, 689–702.

Francis, B. A. (1979). The optimal linear quadratic time invariant regulator with cheap control. *IEEE Trans. Aut. Control*, **AC-24**, 616–621.

Francis, B. A. (1987). *A Course in H∞ Control Theory*. Springer Verlag, Berlin.

Francis, B. A. and W. M. Wonham (1976). The internal model principle of control theory. *Automatica*, **12**, 457–465.

Freudenberg, J. S. and D. P. Looze (1985). Right half plane poles and zeros and design tradeoffs in feedback systems. *IEEE Trans. Aut. Control*, **AC-30**, 555–565.

Glover, K. (1984). All optimal Hankel norm approximations of linear multivariable systems and their L^∞ error bounds. *Int. J. Control*, **39**, 1115–1193.

Grimble, M. J. and M. J. Johnson (1988). *Optimal Control and Stochastic Estimation Theory and Applications*, Vol. 1. John Wiley, New York.

Kwakernaak, H. and R. Sivan (1972a). *Linear Optimal Control Systems*. Wiley Interscience, New York.

Kwakernaak, H. and R. Sivan (1972b). The maximal achievable accuracy of linear optimal regulators and linear optimal filters. *IEEE Trans. Aut. Control*, **AC-17**, 79–86.

Laub, W. S., M. T. Heath, C. C. Paige and R. C. Ward (1987). Computation of system balancing transformation and other applications of simultaneous diagonalization algorithm. *IEEE Trans. Aut. Control*, **AC-32**, 115–122.

Levine, W. S. and M. Athans (1970). On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Trans. Aut. Control*, **AC-15**, 44–48.

MacLean, J. (1990). Modelling and control of a flexible beam. Master's thesis, Dept. of Electrical Eng., Univ. of Toronto, Toronto, Canada.

Miller, D. E. and Davison, E. J. (1989). On necessary assumptions in continuous time model reference adaptive control. In *Proc. IEEE Conf. on Decision and Control*, pp. 1573–1578.

Moore, B. C. (1981). Principle component analysis of linear systems: controllability, observability and model reduction. *IEEE Trans. Aut. Control*, **AC-26**, 17–32.

Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall, Englewood Cliffs, NJ.

Qiu, L. and E. J. Davison (1990). Performance limitations of non minimum phase systems. In *Proc. 11th IFAC World Congress*, Vol. 2, pp. 322–327.

Saberi, A. and P. Sannuti (1987). Cheap and singular control for linear quadratic regulators. *IEEE Trans. Aut. Control*, **AC-32**, 208–219.

Scherzinger, B. M. and E. J. Davison (1985). The optimal LQ regulator with cheap control for not strictly proper systems. *Optimal Control Application and Method*, **6**, 291–303.

Stein, G. and M. Athans (1987). The LQG/LTR procedure for multivariable feedback control design. *IEEE Trans. Aut. Control*, **AC-32**, 105–114.

Tannenbaum, A. (1980). Feedback stabilization of linear dynamical plant with uncertainty in the gain factor. *Int. J. of Control*, **32**, 1–16.

Zhang, Z. and J. S. Freudenberg (1990). Loop transfer recovery for nonminimum phase plants. *IEEE Trans. Aut. Control*, **AC-35**, 547–553.

Automatica, Vol 29, No 2, pp 351–363, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Discrete-time Loop Transfer Recovery for Systems with Nonminimum Phase Zeros and Time Delays*†

ZHIHONG ZHANG‡ and JAMES S. FREUDENBERG§

An analysis of the effects of nonminimum phase zeros and delays on the discrete-time loop transfer recovery procedure leads to a better understanding of the procedure.

Key Words—Discrete time systems, optimal control, nonminimum-phased systems, delays

Abstract—The purpose of this paper is to study what happens when the discrete-time loop transfer recovery (LTR) procedure is applied to plants with nonminimum phase zeros and with uniform time delay in all channels Explicit expressions are given for the asymptotic behavior of the resulting sensitivity function and loop transfer function These results yield a better understanding of the mechanism of the discrete-time loop transfer recovery procedure and the design limitations due to nonminimum phase zeros and time delays

1 INTRODUCTION

SINCE THE SEMINAL work of Kwakernaak (1969) and Doyle and Stein (1979, 1981) the loop transfer recovery method has received a lot of attention and evolved as a formal design procedure (Stein and Athans, 1987). Recently, extension of the loop transfer recovery design technique to discrete-time systems has been studied by a number of researchers (Maciejowski, 1985; Ishihara and Takeda, 1896; Niemann and Sogaard-Andersen, 1988; Kinnaert and Peng, 1990). Motivation for such an interest can be seen as follows. Firstly, guaranteed feedback properties for the discrete-time linear quadratic optimal regulator or Kalman filter do exist (Safonov, 1980; Shaked, 1986; Anderson and Moore, 1990) although they are not as good as in the continuous-time case. Nevertheless, it is desirable to have a method of recovering these properties. Secondly, the loop transfer recovery procedure significantly simplifies the use of the LQG methodology. Knowing that it will be recovered in the LTR procedure, the designer can mainly concentrate on the design of the state feedback loop.

There are two types of observers for discrete-time systems: predicting observers and filtering observers (Franklin and Powell, 1980). The predicting observer is used when there is a significant computation time and the filtering observer is used when the computation time is negligible. The standard state feedback scheme is dual to the predicting observer, but not to the filtering observer. This implies that there is a fundamental difference between the problem of recovering state feedback loop properties at the plant input using a specially tuned Kalman filter (which is a filtering observer) and the problem of recovering the optimal observer loop properties at the plant output using a specially tuned LQ optimal regulator.

Maciejowski (1985) studies the problem of recovering state feedback properties at the plant output. He shows that if the plant is minimum phase and has no time delays and if the cheap control regulator is applied to the filtering observer, then the feedback loop of the observer can be recovered. He also observed that although it is generally impossible to have perfect recovery when the plant is nonminimum phase or when the predicting observer has to be used, a useful degree of recovery is often obtained. An interpretation of this phenomenon in terms of asymptotic eigenvalue locations was also provided in Maciejowski (1985). Due to the fact that sampling often introduces nonmini-

* Received 3 September 1991, revised 22 May 1992, received in final form 4 June 1992 The original version of this paper was not presented at any IFAC meeting This paper was recommended for publication in revised form by Associate Editor R V Patel under the direction of Editor H. Kwakernaak.

† This work was supported in part by National Science Foundation under Grant ECS-8857510 It was conducted while the first author was with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, U S A.

‡ Systems Engineering Center, General Motors Corporation, 1151 Crooks Road, Troy, MI 48084, U S A

§ Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, U.S.A.

mum phase zeros (Åström *et al.*, 1984) that computation time is not always negligible, and that many plants contain time delays, it is important to understand what happens when the loop transfer recovery procedure is applied under these conditions. The problem of loop transfer recovery for a plant with delays has been studied in Kinnaert and Peng (1990) for minimum phase systems. We shall see that their results can be included in our framework as a special case.

In this paper, we shall use the same approach as in Zhang and Freudenberg (1990) to study the effect of nonminimum phase zeros and time delays upon the LTR procedure. In Zhang and Freudenberg (1990) we studied loop transfer recovery for continuous-time nonminimum phase plants. We first factorized the plant into a minimum phase part and an all-pass factor expressed in terms of plant nonminimum phase zeros and their associated directions, and used these factors to derive explicit expressions for the loop transfer function and sensitivity function resulting from the LTR procedure. Design interpretations were then obtained from those expressions. The results of Zhang and Freudenberg (1990) were derived for the problem of recovery at the plant input and could also be applied, via duality, to the problem of recovery at the plant output. Since duality is not so complete with discrete-time systems, we shall only consider here the problem of recovery at the plant output. There are several procedures to achieve the loop transfer recovery (e.g. Niemann and Sogaard-Andersen, 1988; Tsui, 1989; Saberi and Sannuti, 1990; Chen *et al.*, 1991). Our discussion will focus on the one based on Riccati equations (Maciejowski, 1985).

The rest of the paper is organized as follows. Section 2 contains definitions, some properties of nonminimum phase systems and a generalization of the minimum phase/all-pass factorization formulas in Enns (1984) and Zhang and Freudenberg (1990) to discrete-time systems. In Section 3, we derive explicit expressions for the limiting values of the sensitivity and loop transfer functions when the LTR procedure proposed in Maciejowski (1985) is applied to a nonminimum phase plant. Interpretations of the results are also given. The results of applying the LTR procedure with the predicting, rather than the filtering, observer are discussed in Section 4. In Section 5 we study the application of the LTR procedure to a plant with uniform time delay in all channels. An example is given in Section 6 to illustrate our results and conclusions are found in Section 7. An abbreviated version of this paper was presented in Zhang and Freudenberg (1991).

2. NOTATION AND ALL-PASS FACTORIZATION

In the following, we consider a discrete-time system described by state equations

$$x_{k+1} = Ax_k + Bu_k, \tag{1}$$

$$y_k = Cx_k, \tag{2}$$

where $x \in R^n$ is the state, $u \in R^m$ is the control input, and $y \in R^m$ is the measured output. It is assumed that (C, A) is observable, (A, B) is controllable, C and B are each full rank, and $G(z) := C(zI - A)^{-1}B$. We shall use superscripts T and H to denote transpose and complex conjugate transpose, respectively. Notation $\bar{a}$ denotes complex conjugate of a complex number a.

Consider the state feedback control law

$$u_k = -K_c x_k. \tag{3}$$

Let K_c be obtained by using the LQ optimization technique with the performance index

$$J = \sum_{k=0}^{\infty} (u_k^T R u_k + x_k^T Q x_k), \tag{4}$$

where $R > 0$ and $Q = C^T C$. Then it is known that K_c is given by

$$K_c = (R + B^T M B)^{-1} B^T M A, \tag{5}$$

where M is the positive definite solution of the Riccati equation

$$M = A^T M A - A^T M B (R + B^T M B)^{-1} \times B^T M A + Q. \tag{6}$$

Define the loop transfer function of the optimal regulator loop

$$H_{sf}(z) = K_c(zI - A)^{-1}B = (R + B^T M B)^{-1} \times B^T M A (zI - A)^{-1} B. \tag{7}$$

The above control law assumes that all states are available for feedback. Typically, not all the states are measurable, and the missing states must be estimated from the output measurements y_l, $l \le k$. There are two versions of full order estimates for the state x_k: the *filter* estimate $\hat{x}_{k/k}$ which is based on measurements up to and including the current measurement y_k, and the predictor estimate $\hat{x}_{k/k-1}$ which is based on measurements up to y_{k-1}. Two observers commonly used in discrete-time systems are described as follows (Maciejowski, 1985; Franklin and Powell, 1980).

2.1. *Predicting observer*

The predicting observer is described by

$$\hat{x}_{k+1/k} = A\hat{x}_{k/k-1} + Bu_k + K_p(y_k - C\hat{x}_{k/k-1}), \tag{8}$$

where the observer gain K_p is chosen so that $A - K_pC$ is stable. When the predicting observer

is used, the control law (3) is replaced by

$$u_k = -K_c\hat{x}_{k/k-1}. \tag{9}$$

This results in the predicting compensator

$$U(z) = -F_p(z)Y(z), \tag{10}$$

where

$$F_p(z) := K_c[zI - A + BK_c + K_pC]^{-1}K_p. \tag{11}$$

The predicting compensator is appropriate when the computation time is not negligible, and we cannot use the current measurement of the output to update the control.

2.2. *Filtering observer*

The filtering observer is described by

$$\hat{x}_{k+1/k} = A\hat{x}_{k/k-1} + Bu_k + AK_f(y_k - C\hat{x}_{k/k-1}), \tag{12}$$

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + K_f(y_k - C\hat{x}_{k/k-1}), \tag{13}$$

where the observer gain K_f is chosen such that $A - AK_fC$ is stable. In this case, the control law (3) is replaced by

$$u_k = -K_c\hat{x}_{k/k}. \tag{14}$$

This results in the filtering compensator

$$U(z) = -F_f(z)Y(z), \tag{15}$$

where

$$F_f(z) := zK_c[zI - (I - K_fC)(A - BK_c)]^{-1}K_f. \tag{16}$$

The filtering compensator is appropriate when the computation time is negligible, and we can use the current measurement of the output to update the control.

When equation (12) is used alone, we call it the 'predicting' version of the filtering observer, since it has the same form as a predicting observer with gain $K_p = AK_f$.

The Kalman filter is a filtering observer with

$$K_f = PC^T(CPC^T + V)^{-1} \tag{17}$$

where P is the unique positive semidefinite solution to the dual algebraic Riccati equation

$$P = APA^T - APC^T(CPC^T + V)^{-1}CPA^T + W, \tag{18}$$

and $W \geq 0$ and $V > 0$ are fictitious process and measurement noise covariance matrices, respectively. Define the loop transfer function of the Kalman filter loop (see Fig. 1)

$$H_{ob}(z) = C(zI - A)^{-1}AK_f = C(zI - A)^{-1} \times APC^T(CPC^T + V)^{-1}. \tag{19}$$

Note that the optimal state feedback described by (1)–(6) and the optimal observer described by (12)–(13) and (17)–(18) are not dual, but that the optimal state feedback loop transfer function $H_{sf}(z)$ and the optimal observer loop transfer function $H_{ob}(z)$ are dual. Hence, the optimal state feedback is dual to the 'predicting' version of the optimal observer described by (12) and (17)–(18).

Guaranteed feedback properties of the discrete-time optimal regulator and Kalman filter loops also exist (Safonov, 1980; Shaked, 1986; Anderson and Moore, 1990). These properties can be derived from the discrete-time Kalman equality (or its dual)

$$[I + B^T\Phi(z^{-1})^TK_c^T](B^TMB + R)[I + K_c\Phi(z)B] = R + B^T\Phi(z^{-1})^TQ\Phi(z)B, \tag{20}$$

where $\Phi(z) := (zI - A)^{-1}$. These guaranteed properties are not as good as in the continuous case. In particular, there is no infinite gain margin. Of course, this makes sense because the zero at infinity is outside the unit circle and thus in the unstable region.

We now introduce some notation. First, let us define the following transfer function matrices.

Observer loop transfer function: $H(z) = C(zI - A)^{-1}AK_f$.

Observer sensitivity function: $S_{ob}(z) = [I + H(z)]^{-1}$.

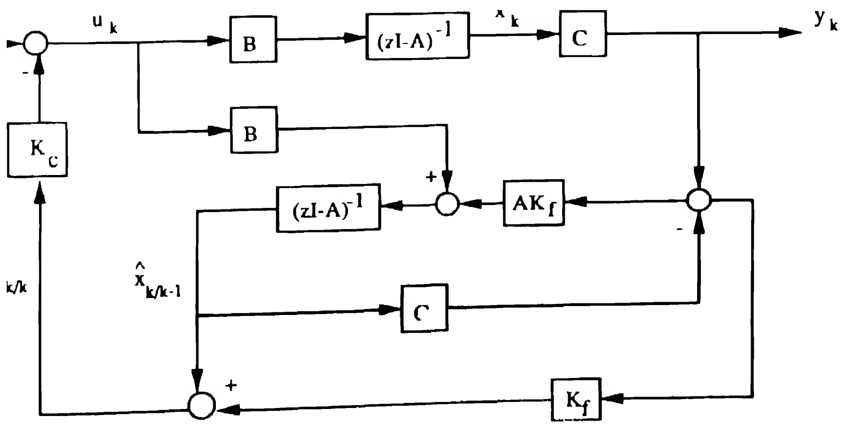


Fig. 1. System structure for the discrete-time loop transfer recovery.

Output feedback loop transfer function: $L(z) = G(z)F(z)$.
Output feedback sensitivity function: $S_{out}(z) = [I + L(z)]^{-1}$.
$F(z) = F_f(z)$ or $F_p(z)$, depending upon which observer-based compensator is used. The following definitions are also needed in the subsequent development.

Definition 2.1 (Davison and Wang, 1974). The transmission zeros of system (1)–(2) are defined to be the set of complex numbers a which satisfy the following inequality

$$\operatorname{rank}\begin{bmatrix} aI - A & -B \\ -C & 0 \end{bmatrix} < n + m. \tag{21}$$

The multiplicity of a is equal to its algebraic multiplicity as defined in MacFarlane and Karcanias (1976).

Definition 2.2 (MacFarlane and Karcanias, 1976). Let a be a transmission zero of $G(z)$, so that

$$[x^T \quad w^T]\begin{bmatrix} aI - A & -B \\ -C & 0 \end{bmatrix} = 0, \tag{22}$$

has a solution with $w^H w = 1$. Then x is called the left state zero direction and w is called the output zero direction.

Definition 2.3. The system (1)–(2) is said to be nonminimum phase if at least one of its transmission zeros is outside the closed unit disk in the complex plane and such zeros are called nonminimum phase zeros of the system. Otherwise the system (1)–(2) is said to be minimum phase.

Note that we call a system minimum phase if it has no finite zeros outside the unit circle.

It is well known that nonminimum phase zeros of a transfer function $G(z)$ may be collected into a stable all-pass factor, i.e. $G(z)$ may be written as

$$G(z) = C_a(z)G_m(z), \tag{23}$$

where $C_a(z)$ is stable, has zeros coinciding with the nonminimum phase zeros of $G(z)$, and satisfies $C_a(z^{-1})^T C_a(z) = I$. The transfer function $G_m(z)$ is minimum phase and is termed the minimum phase counterpart of $G(z)$. The following lemma gives a constructive procedure for performing this factorization.

Lemma 2.1. Given the transfer function $G(z) = C(zI - A)^{-1}B$ that has l nonminimum phase zeros $a_1, a_2, \ldots, a_l$, (including multiplicities), the factorization (23) can be obtained using the following iterative procedure.

Factor out the nonminimum phase zeros of $G(z)$ one at a time as follows:

$$G(z) = C_{a_1}(z)G_m^1(z),$$
$$G_m^1(z) = C_{a_2}(z)G_m^2(z),$$
$$\ldots,$$
$$G_m^{l-1}(z) = C_{a_l}(z)G_m^l(z),$$

where $G_m^i(z) := C_m^i(zI - A)^{-1}B$, $C_m^0 := C$, and for $i = 1, 2, \ldots, l$,

$$C_{a_i}(z) = I - \left(\frac{a_i\bar{a}_i - 1}{a_i + 1}\right)\left(\frac{z + 1}{z\bar{a}_i - 1}\right)\eta_i\eta_i^T, \tag{24}$$

$$C_m^i = C_m^{i-1} - \left(\frac{a_i\bar{a}_i - 1}{\bar{a}_i + 1}\right)\bar{\eta}_i\xi_i^T(A + I). \tag{25}$$

The vectors η_i and ξ_i are solutions of

$$[\xi_i^T \quad \eta_i^T]\begin{bmatrix} a_iI - A & -B \\ -C_m^{i-1} & 0 \end{bmatrix} = 0, \tag{26}$$

with $\eta_i^H\eta_i = 1$. Then the factorization (23) is given by $G_m(z) = G_m^l(z)$ and $C_a(z) = C_{a_1}(z)C_{a_2}(z)\cdots C_{a_l}(z)$.

Proof. Straightforward by applying a bilinear transformation to the continuous-time results of Zhang and Freudenberg (1990). Details can be found in Appendix B of Zhang (1990).

The above procedure yields a formula for the minimum-phase/all-pass factorization which is useful for our purpose. Other expressions for this factorization include the standard inner–outer factorization (e.g. Francis, 1987) and a factorization due to Shaked (1990).

From the above lemma we can see that for a given transfer function $G(z) = C(zI - A)^{-1}B$ there always exist C_m and $C_a(z)$ such that (23) can be written in the following form:

$$G(z) = C_a(z)C_m(zI - A)^{-1}B. \tag{27}$$

We shall now calculate the limiting value of the optimal regulator gain as the control cost approaches zero.

Theorem 2.1. Consider a nonminimum phase system (A, B, C) and its minimum phase counterpart (A, B, C_m), with C_m as calculated in Lemma 2.1. Let K_r be the feedback gain calculated according to (5) and (6) with $R = (1/q^2)I$. Suppose that $CB = CAB = \cdots = CA^{l-2}B = 0$ and $\det(CA^{l-1}B) \neq 0$. Then $K_r \to (C_mA^{l-1}B)^{-1}C_mA^l$ as $q \to \infty$.

Proof. Use duality and Theorem 3.1 of Shaked (1985). Note that since $C_a(z)$ is bicausal, the structure of $G(z)$ at infinity is the same as that of $G_m(z)$ which implies that the conditions $CB = CAB = \cdots = CA^{l-2}B = 0$, $\det(CA^{l-1}B) \neq$

0 are equivalent to the conditions $C_mB = C_mAB = \cdots = C_mA^{l-2}B = 0$, $\det(C_mA^{l-1}B) \neq 0$. Also note that the assumptions in Shaked (1985) that A is stable and that the system (A, B, C) has no poles or zeros at the origin can be removed (for details, see Appendix A of Zhang (1990)).

3. LOOP TRANSFER RECOVERY WITH NONMINIMUM PHASE ZEROS

The development of this section is based on the discrete-time loop transfer recovery procedure proposed by Maciejowski (1985). We first briefly state the procedure. Consider the observer-based output feedback system shown in Fig. 1. Suppose that the plant transfer matrix is minimum phase and $\det(CB) \neq 0$. The procedure is to design the observer loop $H(z) = C(zI - A)^{-1}AK_f$ to meet design specifications (with augmented dynamics if necessary) and then to recover this loop asymptotically by tuning the state feedback gain K_c which is obtained from (5) and (6) with $R = (1/q^2)I$. If the filtering compensator $F_f(z)$ is used, then perfect recovery can be obtained asymptotically at the plant output, i.e. as $q \to \infty$,

$$G(z)F_f(z) \to H(z), \tag{28}$$

$$S_{out}(z) \to S_{ob}(z). \tag{29}$$

Here, as in the sequel, convergence of transfer functions is pointwise in z.

There are three crucial assumptions in the above LTR procedure, namely, the requirements that the plant is minimum phase, that $\det(CB) \neq 0$, and that the filtering compensator is used. We study next what happens when the minimum phase requirement is not satisfied. The use of the predicting compensator and the case where $\det(CB) = 0$ will be discussed in the subsequent sections. The following theorem reveals the asymptotic behavior of the loop transfer and sensitivity functions when the plant is nonminimum phase.

Theorem 3.1. Suppose the plant $G(z)$ is factored as in (27), $\det(CB) \neq 0$, and the LTR procedure is applied using the filtering compensator $F_f(z)$. Then, as $q \to \infty$, the asymptotic behavior of the filtering compensator is given by

$$\begin{aligned} F_f(z) \to\ & (C_m\Phi(z)B)^{-1} \\ & \times [I - H_m(z)(I + H(z))^{-1}C_a(z)]^{-1} \\ & \times H_m(z)(I + H(z))^{-1}, \end{aligned} \tag{30}$$

where $H_m(z) := C_m(zI - A)^{-1}AK_f$. Consequently, the asymptotic values of the loop transfer and the sensitivity functions are given by

$$L(z) \to [H(z) - E(z)][I + E(z)]^{-1}, \tag{31}$$

$$S_{out}(z) \to [I + E(z)]S_{ob}(z), \tag{32}$$

where $E(z) := [C - C_a(z)C_m](zI - A)^{-1}AK_f$ is called the error function.

Proof. See Appendix A.

Clearly, when the plant is minimum phase, (31) and (32) reduce to the known results (28) and (29), and the recovery is seen to be perfect. If the plant is nonminimum phase, on the other hand, then the quality of recovery at a given frequency depends upon the size of the error function $E(z)$ at that frequency. An inspection of the error function reveals that if, in addition to the assumptions of Theorem 3.1, the observer loop has the same nonminimum phase zero structure as the plant, i.e. if

$$H(z) = C_a(z)C_m(zI - A)^{-1}AK_f,$$

then perfect recovery can be obtained. Also one can see from (30) that, in the minimum phase, case, the compensator constructs an inverse of the plant and substitutes the observer loop and that, in the nonminimum phase case, an inverse of the minimum phase counterpart of the plant is constructed. This provides a clear picture of how the recovery process works and complements the pole–zero cancellation explanation given by Maciejowski (1985). Next, we give a formula for calculating the error function $E(z)$ in terms of the nonminimum phase zeros and their associated directions.

Lemma 3.1. Let $G(z)$ have l nonminimum phase zeros $a_1, a_2, \ldots, a_l$. Define $C_a^0(z) = I$, $C_a^k(z) = C_{a_1}(z)C_{a_2}(z)\cdots C_{a_k}(z)$, $E^0(z) = 0$, and $E^k(z) = (C - C_a^k(z)C_m^k)(zI - A)^{-1}AK_f$ for $k = 1, 2, \ldots, l$. Then

$$\begin{aligned} E(z) = E^l(z) = \sum_{k=1}^{l} & \frac{a_k\bar{a}_k - 1}{z\bar{a}_k - 1} \\ & \times C_a^{k-1}(z)\bar{\eta}_k\xi_k^TAK_f, \end{aligned} \tag{33}$$

with C_m^k and $C_{a_k}(z)$ defined by (24) and (25), and ξ_k and η_k defined by (26).

Proof. See Appendix B.

The following theorem states that perfect recovery can in fact be obtained in output directions that are orthogonal to those associated with the nonminimum phase zeros.

Theorem 3.2. Suppose that plant $G(z)$ has l distinct nonminimum phase zeros $a_1, a_2, \ldots, a_l$ and that $\det(CB) \neq 0$. Let $w_1, w_2, \ldots, w_l$ denote the corresponding output zero directions defined by (22). Define W to be the subspace of

C^m spanned by $w_1, w_2, \ldots, w_l$, and $W^\perp$ to be its orthogonal complement in C^m. Let $P_{W^\perp}$ be a projection onto $W^\perp$. Then if the LTR procedure is applied using the filtering compensator $F_f(z)$, we have

$$P_{W^\perp}^T S_{out}(z) \rightarrow P_{W^\perp}^T S_{ob}(z), \tag{34}$$

as $q \rightarrow \infty$.

Proof. See Appendix C.

An analogous result may be obtained for the dual version of the continuous-time recovery result in Zhang and Freudenberg (1990).

For systems that have a single (real) nonminimum phase zero, more insightful expressions can be obtained.

Corollary 3.1. Consider a plant $G(z)$ that has only one nonminimum phase zero, and assume it is at $z = a$. Suppose that $\det(CB) \neq 0$ and that the LTR procedure is applied using the filtering compensator $F_f(z)$. Then the sensitivity function of the system satisfies

$$S_{out}(z) \rightarrow \left[I + \frac{a^2 - 1}{za - 1} ww^T H(a)\right] S_{ob}(z), \tag{35}$$

as $q \rightarrow \infty$, where $w^T H(a) = w^T C(aI - A)^{-1} AK_f$ and w is the output zero direction determined by (22). If the plant is scalar, then

$$S_{out}(z) \rightarrow \left[1 + \frac{a^2 - 1}{sa - 1} H(a)\right] S_{ob}(z), \tag{36}$$

as $q \rightarrow \infty$.

Proof. It follows from Lemma 3.1 by setting $l = 1$ that

$$E(z) = \frac{a^2 - 1}{za - 1} \eta_1 \xi_1^T AK_f.$$

The limit (35) follows by noting $\xi_1^T AK_f = \eta_1^T C(aI - A)^{-1} AK_f$ and $\eta_1 = w$. For scalar systems, $w = 1$. Hence (35) reduces to (36).

One can see that if the target feedback loop $H(z)$ also has a zero at $z = a$, with output zero direction w, then perfect recovery is possible. Otherwise there will be an unavoidable error in recovery, whose size depends upon the value of $\|w^T H(a)\|$, and thus upon the location of the zero relative to the frequency range over which the target feedback loop gain is large in the direction w. It is clear from the above results that if the nonminimum phase zero is far outside the bandwidth of the target feedback loop, then good recovery can be obtained. This confirms the observation by Maciejowski (1985). Also, there appears to be a tradeoff between the feedback properties of the target feedback loop and the quality of recovery. Further quantifications of this tradeoff may be performed using Poisson integral relations similarly to the continuous case (Zhang and Freudenberg, 1990). This has recently been done by Léon de la Barra (1991).

4. USE OF THE PREDICTING COMPENSATOR

The recovery procedure in the previous section assumes negligible computation time, so that it is possible to implement the filtering compensator. This may be impractical, since the time required to compute the control signal is not always negligible. If this is the case, then the feedback law (9) has to be used, resulting in the predicting compensator $F_p(z)$ with $K_p = AK_f$. Although it has been shown (Maciejowski, 1985) that in this case perfect recovery cannot generally be obtained, it is of interest to investigate what happens if we try to apply the recovery procedure using the predicting compensator.

Theorem 4.1. Suppose that the plant $G(z)$ is factored as in (27), that $\det(CB) \neq 0$, and that the LTR procedure is applied using the predicting compensator $F_p(z)$ with $K_p = AK_f$. Then, as $q \rightarrow \infty$, the asymptotic values of the loop transfer and the sensitivity functions are given by

$$L(z) \rightarrow [H(z) - E_p(z)][I + E_p(z)]^{-1}, \tag{37}$$

$$S_{out}(z) \rightarrow [I + E_p(z)] S_{ob}(z), \tag{38}$$

where the error function is given by

$$E_p(z) := z^{-1}[zC - C_a(z)C_m A] \times (zI - A)^{-1} AK_f. \tag{39}$$

Proof. Similar to proof of Theorem 3.1, straightforward calculation of $G(z)F_p(z)$ by substituting $(C_m B)^{-1} C_m A$ for K_c.

The error function $E_p(z)$ can be calculated using the formulas for $C_a(z)$ and C_m that we developed in Section 2. Note that the above result still holds without assuming $K_p = AK_f$, as long as the target observer loop $H(z)$ is interpreted as $C(zI - A)^{-1} K_p$ in such a case.

Corollary 4.1. Suppose, in addition to the assumptions of Theorem 4.1, that the plant $G(z)$ is minimum phase. Then the error function $E_p(z)$ reduces to

$$E_p(z) = z^{-1} CAK_f. \tag{40}$$

It follows that if the observer gain satisfies

$$CAK_f = 0, \tag{41}$$

then perfect recovery can be obtained.

As noted by Maciejowski (1985), perfect recovery is generally unattainable with the predicting compensator even for minimum phase plants. The difference between the results obtained in this case for the two compensators $F_f(z)$ and $F_p(z)$ can be explained as follows. Note first that the assumption $\det(CB) \neq 0$ implies that the plant has an inherent one-step delay in all channels. Suppose that the observer loop also has a one-step delay. Then the only way that this loop can be perfectly recovered by the loop with observer-based compensator is for this compensator to have a direct feedthrough from the plant output to the control input, i.e. to have a proper inverse. An inspection of (11), (16), and Fig. 1 reveals that this is possible only if the filtering compensator is implemented. On the other hand, suppose that the observer gain satisfies (41), so that the observer loop has (at least) an inherent two-step delay. Then it becomes potentially possible for the observer loop to be recovered using the predicting compensator. Unfortunately, requiring such a two-step delay would result in a less satisfactory observer loop than would be the case if a shorter delay were present. This situation is analogous to that studied in the previous section, where we saw that only nonminimum phase (and therefore inferior) observer loops could be recovered. Hence, we see that when the computation time is not negligible and the predicting compensator has to be used, perfect recovery is possible only if the target observer loop is constrained to have an extra step delay. This observation is consistent with that of Ishihara and Takeda (1986).

It is of interest to study how recovery takes place when the filtering compensator is implemented to recover an observer loop with (at least) a two-step delay in all channels. From the above discussion, we know that such an observer loop satisfies $CAK_f = 0$. Since the asymptotic control gain satisfies $K_c \to (CB)^{-1}CA$, it follows that the condition $CAK_f = 0$ implies that $K_c K_f \to 0$. Therefore, as $q \to \infty$, the feedback control law

$$u_k = -K_c\hat{x}_{k/k} = -K_c[\hat{x}_{k/k-1} + K_f(y_k - C\hat{x}_{k/k-1})] \to -K_c\hat{x}_{k/k-1}.$$

Hence, asymptotically the optimal state feedback control law does not utilize those states that are updated using the current output measurement. Therefore the direct feedthrough link in Fig. 1 is not used.

The role of the assumption that $\det(CB) \neq 0$ is also now clear. If it is not satisfied, then the plant would have at least a two-step delay in some channels, and it would be impossible for the observer-based output feedback loop to asymptotically recover an observer loop with a one-step delay in all channels.

5. LOOP TRANSFER RECOVERY WITH TIME DELAYS

In the previous sections, the assumption is made that $\det(CB) \neq 0$. However, this assumption will be violated for systems with at least a two-step time delay in some channel of the plant. It has been observed (Maciejowski, 1985) that perfect recovery cannot generally be obtained in this case, but it would be of interest to see what happens when the LTR procedure is applied anyway. In the following, we consider application of the LTR procedure to a class of plants whose delay structure is characterized by

$$CB = CAB = \cdots = CA^{l-2}B = 0,\ \det(CA^{l-1}B) \neq 0, \tag{42}$$

where $l \geq 2$, i.e. plants that have a uniform l-step delay in all channels. This problem has been studied by Kinnaert and Peng (1990) for minimum phase systems using the predicting compensator in the LTR procedure. The following results extend and complement those of Kinnaert and Peng (1990).

Theorem 5.1. Suppose that the plant $G(z)$ is factored as in (27) and has a uniform l-step delay as characterized by (42). Assume that the LTR procedure is applied to the system using either the filtering compensator $F_f(z)$ or the predicting compensator $F_p(z)$ with $K_p = AK_f$. Then, as $q \to \infty$, the asymptotic values of the loop transfer and the sensitivity functions are given by

$$L(z) \to [H(z) - E_d(z)][I + E_d(z)]^{-1}, \tag{43}$$

$$S_{out}(z) \to [I + E_d(z)]S_{ob}(z), \tag{44}$$

where $H(z) = C(zI - A)^{-1}AK_f$. The error function $E_d(z)$ is given by

$$E_d(z) = \left(C - \frac{1}{z^{l-1}}C_a(z)C_mA^{l-1}\right) \times (zI - A)^{-1}AK_f, \tag{45}$$

when the filtering compensator is used, and is given by

$$E_d(z) = \left(C - \frac{1}{z^l}C_a(z)C_mA^l\right) \times (zI - A)^{-1}K_p, \tag{46}$$

when the predicting compensator with $K_p = AK_f$ is used.

Proof. Direct calculation of the loop transfer function using the cheap control gain given in Theorem 2.1.

Notice that (46) also holds for a general predicting compensator with $K_p \neq AK_f$. However, the definition of $H(z)$ in (43) and (44) has to be changed to $H(z) = C(zI - A)^{-1}K_p$ in such a case. If the plant is minimum phase, then (46) reduces to

$$E_d(z) = H(z) - \frac{1}{z^l} CA^l(zI - A)^{-1}K_p.$$

Substituting the above into (43) results in the following limiting value of the loop transfer function

$$\begin{aligned} L(z) &= \frac{1}{z^l} CA^l(zI - A)^{-1}K_p[I + H(z) \\ &\quad - \frac{1}{z^l} CA^l(zI - A)^{-1}K_p]^{-1} \\ &= C(zI - A)^{-1}A^lK_p(z^lI + z^{l-1}CK_p \\ &\quad + \cdots + CA^{l-1}K_p)^{-1}, \end{aligned}$$

which gives the same result as Theorem 3 of Kinnaert and Peng (1990).

One can see from Theorem 5.1 that the recovery error $E_d(z)$ is a function of the nonminimum zeros, the time delays, and the observer loop gain (K_f or K_p) which implies that all those factors can affect the quality of recovery. For a given target observer loop, the recovery error can be calculated, *a priori*, using (45) or (46) and our formulas for $C_a(z)$ and C_m, to determine how much and at which frequencies it can be recovered. Since $E_d(z)$ is generally nonzero, perfect recovery cannot be obtained in general and the quality of recovery at a certain frequency will depend upon the size of the error function at that frequency. However, if the observer loop to be recovered meets certain constraints, then perfect recovery is possible. First, we consider the case where the filtering compensator is used. It follows from (45) that if the observer loop transfer function $H(z) = C(zI - A)^{-1}AK_f$, i.e. the transfer function to be recovered, satisfies

(1) $H(z) = C_a(z)C_m(zI - A)^{-1}AK_f$,

(2) $CAK_f = CA^2K_f = \cdots = CA^{l-1}K_f = 0$,

then the error function $E_d(z)$ will be identically zero. In other words, if the observer loop is chosen to have the same nonminimum phase structure and at least as many steps of time delay as the plant, then it can be perfectly recovered using the filtering compensator.

For the case where the general predicting compensator is used, if the observer loop $H(z) = C(zI - A)^{-1}K_p$ satisfies

(1) $H(z) = C_a(z)C_m(zI - A)^{-1}K_p$,

(2) $CK_p = CAK_p = \cdots = CA^{l-1}K_p = 0$,

then one can show that

$$H(z) = \frac{1}{z^l} C_a(z)C_mA^l(zI - A)^{-1}K_p,$$

which, by (46), implies that the recovery error $E_d(z)$ is zero. Hence, to have perfect recovery using the predicting compensator, the target observer loop has to have the same nonminimum phase zero structure as the plant and have at least one more step of time delay than the plant.

From the above discussion, one can see that the nonminimum phase plant zeros, plant time delays, and controller computation delays all impose constraints upon the class of recoverable target loop transfer functions. Since the recovery error is a function of the total time delay in the feedback loop, it follows that time delays in the plant and controller have the same effect on the loop transfer recovery. This observation is consistent with that of Kinnaert and Peng (1990).

Finally, note that our results concerning LTR using the predicting compensator could also be applied, via duality, to the problem of recovery at the plant input.

6. EXAMPLE

To illustrate some of our results, let us apply the LTR procedure to a nonminimum phase plant. Consider a sampled data system composed of a zero-order hold, a linear time-invariant continuous system $G(s) = \frac{1}{(s+1)^3}$ and a sampler in series (see Fig. 2).

The corresponding pulse transfer function with sampling period T is given by Åström *et al.* (1984).

$$G(z) = \frac{b_1z^2 + b_2z + b_3}{(z - e^{-T})^3}$$

where

$$b_1 = 1 - \left(1 + T + \frac{T^2}{2}\right)e$$

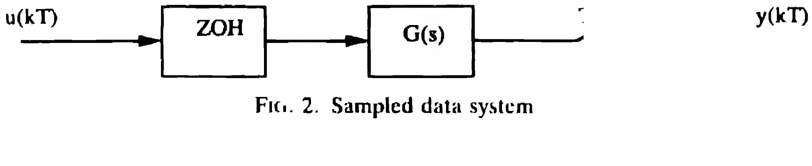


FIG. 2. Sampled data system

$$b_2 = \left(-2 + T + \frac{T^2}{2}\right)e^{-T} + \left(2 + T - \frac{T^2}{2}\right)e^{-2T},$$

$$b_3 = \left(1 - T + \frac{T^2}{2}\right)e^{-2T} - e^{-3T}.$$

This transfer function has a real nonminimum phase zero for $0 < T < 1.8399$. According to Åström *et al.* (1984), nonminimum phase sampling zeros usually lie near the negative real axis when the sampling period is sufficiently small. For this example, the nonminimum phase zero approaches -3.732 as the sampling period goes to zero. It follows from Corollary 3.1 that good recovery may be obtained for this example when the sampling period is small, since the nonminimum phase sampling zero is expected to be outside the bandwidth of the target observer loop. We shall see in the following that this is indeed the case.

A realization of $G(z)$ is given by

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ e^{-3T} & -3e^{-2T} & 3e^{-T} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$$C = [b_3 \quad b_2 \quad b_1].$$

The observer loop is designed using (17)–(18) with weightings $V = 1$ and $W = BB'$. Results obtained by applying the LTR procedure with the filtering compensator are given in Figs 3–9. For the first set of figures (Figs 3–5), we choose the sampling period $T = 0.05$ which results in a nonminimum phase zero at $z = -3.5949$. The plots demonstrate that the sensitivity function $S_{out}(z)$ indeed converges to the function we predicted in Theorem 3.1. The second set of figures (Figs 6–9) shows how the length of the sampling period affects the location of the nonminimum phase zero $z = a$ which, in turn, affects the quality of recovery. First, we notice that the nonminimum phase sampling zero moves away from the unit disk along the real negative axis as the sampling period decreases. This implies that the nonminimum phase sampling zero gets farther away from the bandwidth of the target loop transfer function as the sampling period gets smaller. In these figures we plot the recoverable sensitivity function (solid line) and the desired target sensitivity function (dashed line). As expected, we observe that the recoverable sensitivity function approaches the target sensitivity function as the sampling period decreases.

7. CONCLUSIONS

In this paper, we have studied applications of the discrete-time LTR procedure to plants with nonminimum phase zeros and time delays. Explicit expressions are derived for the asymptotic behavior of the sensitivity function and loop transfer function resulting from the LTR procedure. The results are given for both filtering compensator and predicting compensator cases. For a given target loop, these expressions show *a priori*, how much, at which frequencies and in what directions the loop can be recovered. From our results, we can see that the LTR procedure, if used properly, may still be an effective design approach for discrete-time systems with nonminimum phase zeros and time delays. The plant nonminimum phase zero and time delay structures and computation delays

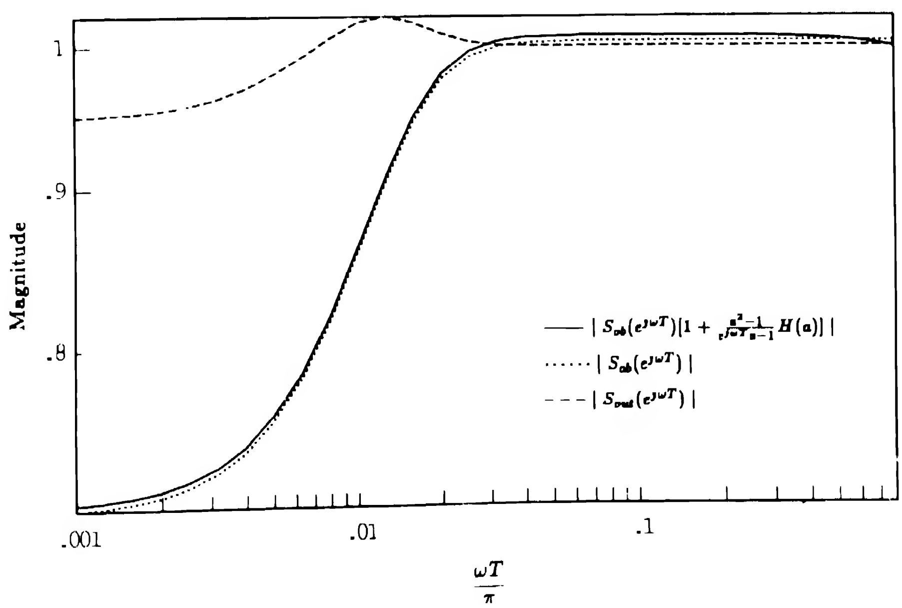


FIG. 3. Discrete-time asymptotic recovery: $q = 1$.

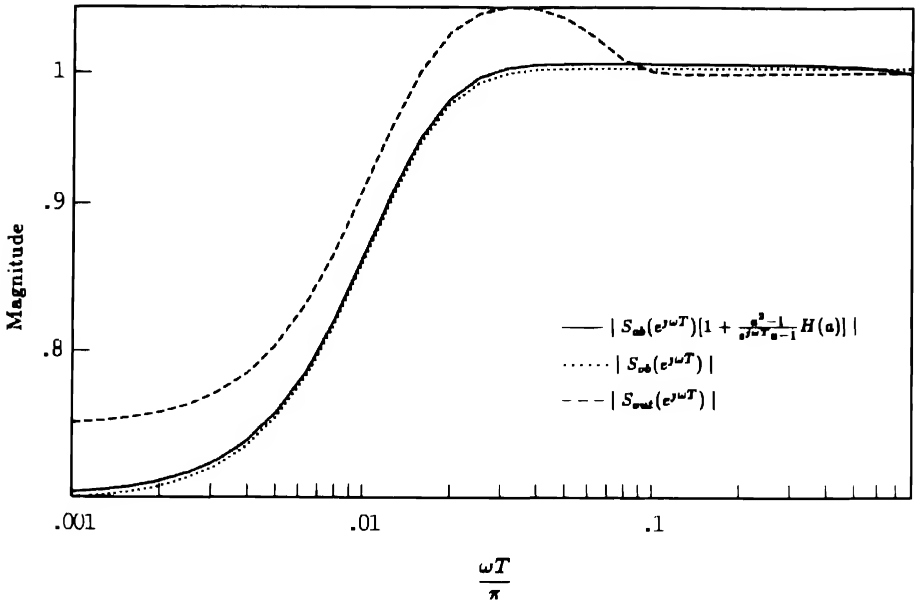


FIG. 4 Discrete-time asymptotic recovery: $q = 100$.

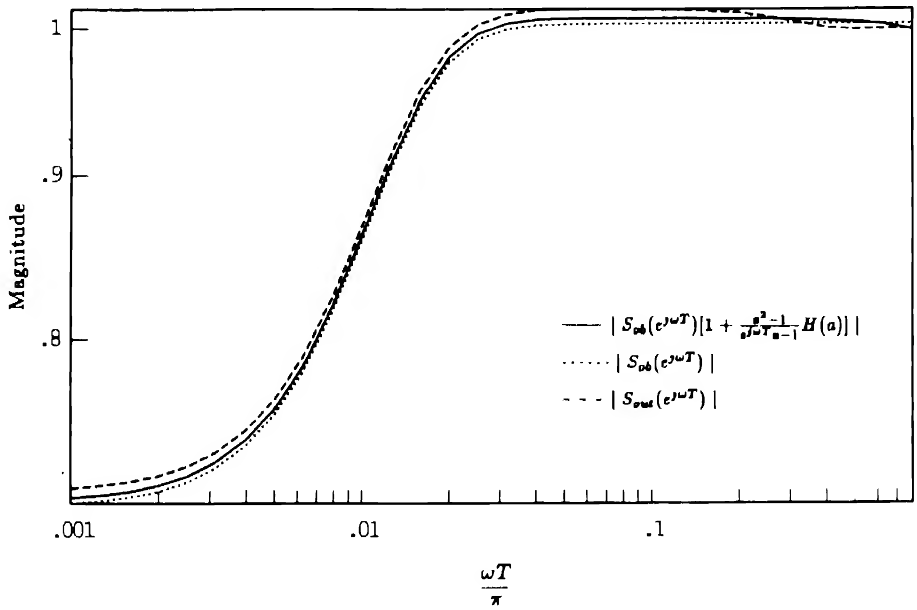


FIG. 5. Discrete-time asymptotic recovery: $q = 10{,}000$

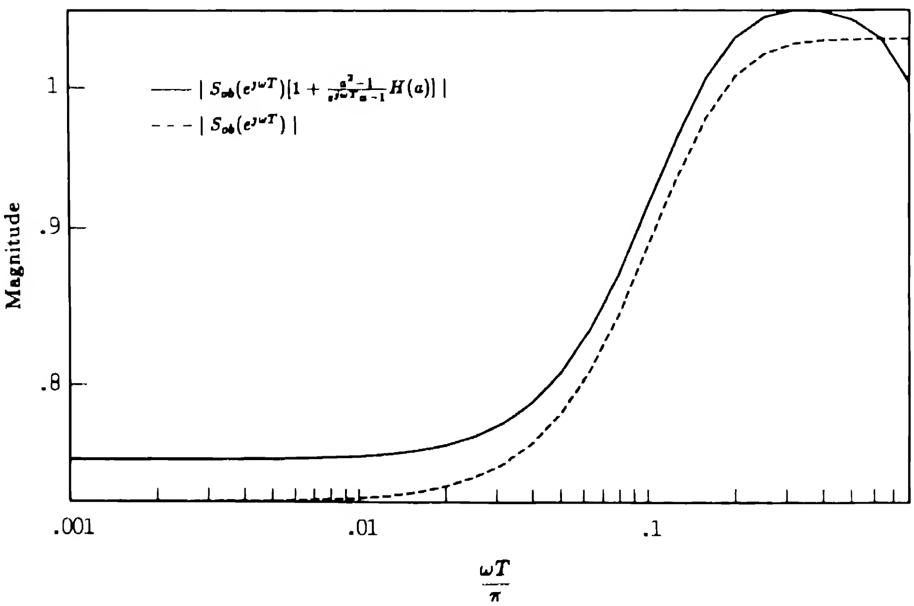


FIG. 6. Discrete-time LTR vs sampling period: $T = 0.5$ and $a = -2.5782$.

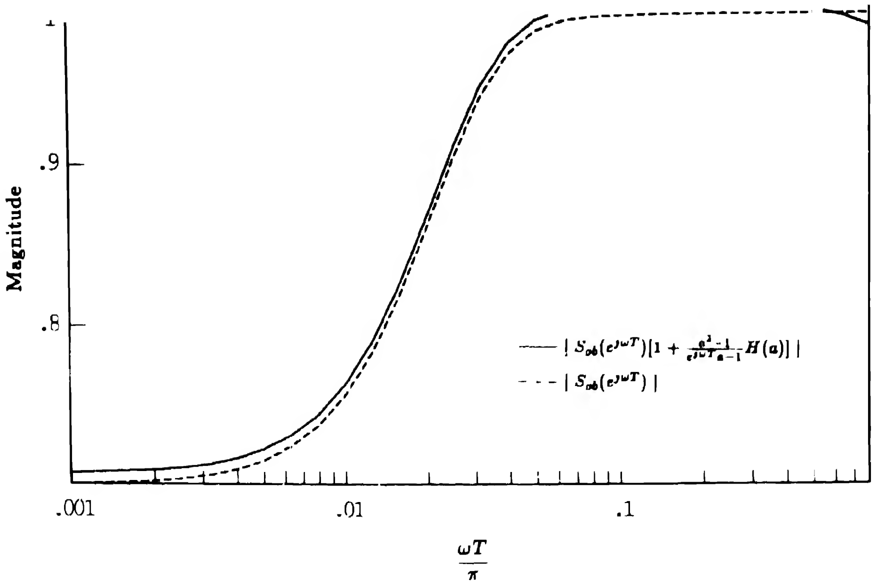


FIG. 7. Discrete-time LTR vs sampling period: $T = 0.1$ and $a = -3.4631$.

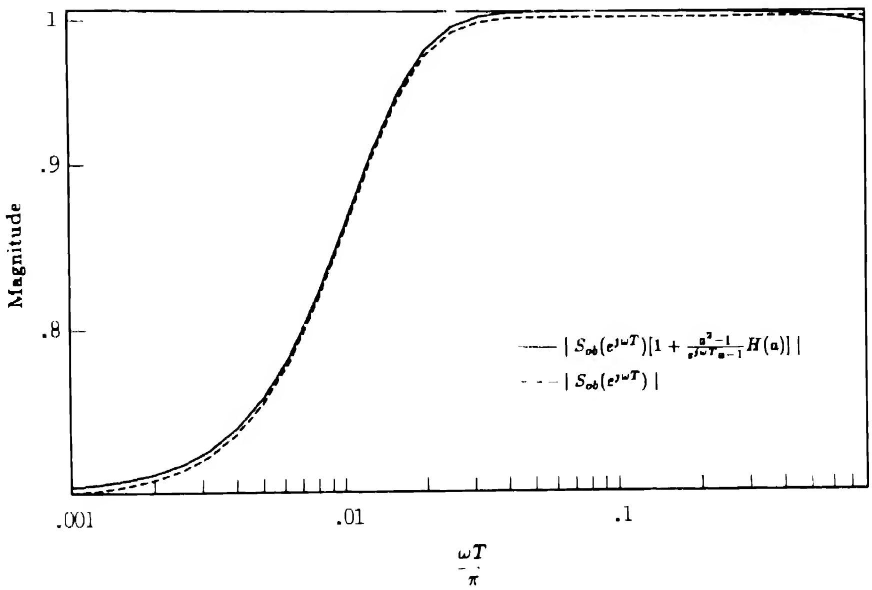


FIG. 8. Discrete-time LTR vs sampling period: $T = 0.05$ and $a = -3.5949$.

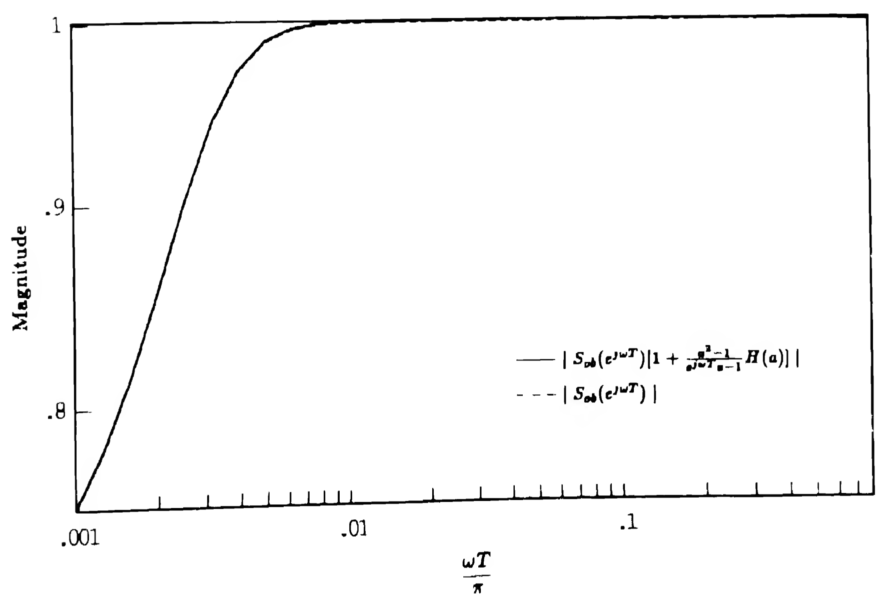


FIG. 9. Discrete-time LTR vs sampling period: $T = 0.01$ and $a = -3.7042$.

essentially impose certain constraints upon the recoverable target loop transfer functions. In this paper, it is assumed that the plant is square and has uniform delay in all channels; further research is needed to extend the results to cases where those assumptions are not satisfied. The general results on singular discrete-time filtering problem by Shaked (1985) may prove useful in this regard.

REFERENCES

Anderson, B. and J. B. Moore (1990). *Optimal Control: Linear Quadratic Methods*. Prentice Hall, Englewood Cliffs, NJ.

Åström, K. J., P. Hagander and J. Sternby (1984). Zeros of sampled systems. *Automatica*, **20**, 31–38.

Chen, B. M., A. Saberi and P. Sannuti (1991). A new stable compensator design for exact and approximate loop transfer recovery. *Automatica*, **27**, 257–280.

Davison, E. J. and S. H. Wang (1974). Properties and calculation of transmission zeros of linear multivariable systems. *Automatica*, **10**, 643–658.

Doyle, J. C. and G. Stein (1979). Robustness with observers. *IEEE Trans. Aut. Control*, **AC-24**, 607–611.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: Concepts for a classical/modern synthesis. *IEEE Trans. Aut. Control*, **AC-26**, 4–16.

Enns, D. (1984). Model reduction for control system design. Ph.D. dissertation, Stanford University, Stanford, CA.

Francis, B. A. (1987). *A Course in H_∞ Control Theory*. Springer-Verlag, New York.

Franklin, G. and J. D. Powell (1980). *Digital Control of Dynamic Systems*. Addison-Wesley, New York.

Ishihara, T. and H. Takeda (1986). Loop transfer recovery techniques for discrete-time optimal regulators using prediction estimators. *IEEE Trans. Aut. Control*, **AC-31**, 1149–1151.

Kinnaert, M. and Y. Peng (1990). Discrete-time LQG/LTR techniques for systems with time delays. *Systems and Control Letters*, **15**, 303–311.

Kwakernaak, H. (1969). Optimal low-sensitivity linear feedback systems. *Automatica*, **5**, 279–286.

Léon de la Barra, B. A. (1991). Frequency domain tradeoffs in loop transfer recovery for multivariable nonminimum phase discrete-time systems. Technical Report EE9117, University of Newcastle, Australia.

MacFarlane, A. G. J. and N. Karcanias (1976). Poles and zeros of linear multivariable systems: a survey of the algebraic, geometric and complex-variable theory. *Int. J. Control*, **24**, 33–74.

Maciejowski, J. M. (1985). Asymptotic recovery for discrete-time systems. *IEEE Trans. Aut. Control*, **AC-30**, 602–605.

Niemann, H. H. and P. Sogaard-Andersen (1988). New results in discrete-time loop transfer recovery. *Proc. of American Control Conf.*, pp. 2483–2489.

Saberi, A. and P. Sannuti (1990). Observer design for loop transfer recovery and for uncertain dynamical systems. *IEEE Trans. Aut. Control*, **35**, 878–897.

Safonov, M. G. (1980). *Stability and Robustness of Multivariable Feedback Systems*. MIT Press, MA.

Shaked, U. (1985). Explicit solution to the singular discrete-time stationary linear filtering problem. *IEEE Trans. Aut. Control*, **30**, 34–47.

Shaked, U. (1986). Guaranteed stability margin for the discrete-time linear quadratic optimal regulator. *IEEE Trans. Aut. Control*, **AC-31**, 162–165.

Shaked, U. (1990). An explicit expression for the minimum-phase image of transfer function matrices. *IEEE Trans. Aut. Control*, **34**, 1290–1293.

Stein, G. and M. Athans (1987). The LQG/LTR procedure for multivariable feedback control design. *IEEE Trans. Aut. Control*, **AC-32**, 105–114.

Tsui, C. C. (1989). On loop transfer recovery. *Proc. of American Contr. Conf.*, Pittsburgh, PA.

Zhang, Z. (1990). Loop transfer recovery for nonminimum phase plants and ill-conditioned plants. Ph.D. Thesis, The University of Michigan, Ann Arbor, MI.

Zhang, Z. and J. S. Freudenberg (1990). Loop transfer recovery for nonminimum phase plant. *IEEE Trans. Aut. Control*, **AC-35**, 547–553.

Zhang, Z. and J. S. Freudenberg (1991). On discrete-time loop transfer recovery. *Proc. of American Contr. Conf.*, Boston, MA.

APPENDIX A: PROOF OF THEOREM 3.1.

By Theorem 2.1, we know that $K_c \to (C_m B)^{-1} C_m A$ as $q \to \infty$. Now evaluate the open loop transfer function $L(z)$ in the limit as $q \to \infty$. Define $\Phi(z) = (zI - A)^{-1}$. $F_f(z)$ can be rewritten as

$$\begin{aligned} F_f(z) &= zK_c[I + \Phi(z)BK_c + \Phi(z)K_fC(A - BK_c)]^{-1}\Phi(z)K_f \\ &= zK_c[I + (I + \Phi(z)K_fCA)^{-1}\Phi(z)BK_c \\ &\quad - (I + \Phi(z)K_fCA)^{-1}\Phi(z)K_fCBK_c]^{-1} \\ &\quad \times (I + \Phi(z)K_fCA)^{-1}\Phi(z)K_f. \end{aligned}$$

Note that $H_m(z) = C_m\Phi(z)AK_f$. Since $K_c \to (C_mB)^{-1}C_mA$ as $q \to \infty$, it follows that

$$\begin{aligned} F_f(z) &\to z(C_mB)^{-1}C_mA[I + (I + \Phi(z)K_fCA)^{-1}\Phi(z)B \\ &\quad \times (C_mB)^{-1}C_mA - (I + \Phi(z)K_fCA)^{-1}\Phi(z)K_fCB \\ &\quad \times (C_mB)^{-1}C_mA]^{-1}(I + \Phi(z)K_fCA)^{-1}\Phi(z)K_f \\ &= z(C_mB)^{-1}[I + C_mA(I + \Phi(z)K_fCA)^{-1}\Phi(z)B(C_mB)^{-1} \\ &\quad - H_m(z)(I + H(z))^{-1}CB(C_mB)^{-1}]^{-1}H_m(z)(I + H(z))^{-1} \\ &= z[C_mB + C_mA(I - \Phi(z)K_fCA(I + \Phi(z)K_fCA)^{-1})\Phi(z)B \\ &\quad - H_m(z)(I + H(z))^{-1}CB]^{-1}H_m(z)(I + H(z))^{-1} \\ &= z[C_mB + C_mA\Phi(z)B - H_m(z)(I + H(z))^{-1}CA\Phi(z)B \\ &\quad - H_m(z)(I + H(z))^{-1}CB]^{-1}H_m(z)(I + H(z))^{-1} \\ &= z[C_m(I + A\Phi(z))B - H_m(z)(I + H(z))^{-1} \\ &\quad \times C(I + A\Phi(z))B]^{-1}H_m(z)(I + H(z))^{-1} \\ &= [C_m\Phi(z)B - H_m(z)(I + H(z))^{-1}C\Phi(z)B]^{-1} \\ &\quad \times H_m(z)(I + H(z))^{-1} \end{aligned}$$

Note that $C\Phi(z)B = C_a(z)C_m\Phi(z)B$, we obtain that

$$\begin{aligned} F_f(z) \to (C_m\Phi(z)B)^{-1}[I - H_m(z)(I + H(z))^{-1}C_a(z)]^{-1} \\ \times H_m(z)(I + H(z))^{-1}. \end{aligned}$$

This completes the proof of (30). By definition of $E(z)$, we have that

$$\begin{aligned} L(z) &\to C\Phi(z)B(C_m\Phi(z)B)^{-1}[I - H_m(z)(I + H(z))^{-1} \\ &\quad \times C_a(z)]^{-1}H_m(z)(I + H(z))^{-1} \\ &= C_a(z)H_m(z)[I + H(z) - C_a(z)H_m(z)]^{-1} \\ &= [H(z) - E(z)][I + E(z)]^{-1}, \end{aligned}$$

as $q \to \infty$. This gives (31). Since $S_{out}(z) = [I + L(z)]^{-1}$, (32) follows readily.

APPENDIX B: PROOF OF LEMMA 3.1

First, let us derive a recursive formula for $E^k(z)$. Using the factorization formulas (24) and (25), we get

$$\begin{aligned} &C - C_a^k(z)C_m^k \\ &= C - C_a^{k-1}(z)\left[I - \left(\frac{a_k\bar{a}_k - 1}{a_k + 1}\right)\left(\frac{z + 1}{z\bar{a}_k - 1}\right)\bar{\eta}_k\eta_k^T\right] \\ &\quad \times \left[C_m^{k-1} - \left(\frac{a_k\bar{a}_k - 1}{\bar{a}_k + 1}\right)\bar{\eta}_k\xi_k^T(A + I)\right] \\ &= (C - C_a^{k-1}(z)C_m^{k-1}) + C_a^{k-1}(z)\left[\left(\frac{a_k\bar{a}_k - 1}{a_k + 1}\right)\right. \end{aligned}$$

$$\times\left(\frac{z+1}{z\bar{a}_k-1}\right)\bar{\eta}_k\eta_k^T C_m^{k-1}+\left(\frac{a_k\bar{a}_k-1}{\bar{a}_k+1}\right)\bar{\eta}_k\xi_k^T(A+I)$$
$$-\left(\frac{a_k\bar{a}_k-1}{a_k+1}\right)\left(\frac{z+1}{z\bar{a}_k-1}\right)\left(\frac{a_k\bar{a}_k-1}{\bar{a}_k+1}\right)\bar{\eta}_k\xi_k^T(A+I)\Big]$$
$$=(C-C_a^{k-1}(z)C_m^{k-1})+\frac{(a_k a_k-1)}{(z\bar{a}_k-1)(\bar{a}_k+1)(a_k+1)}C_a^{k-1}(z)$$
$$\times[(z+1)(\bar{a}_k+1)\bar{\eta}_k\eta_k^T C_m^{k-1}+(za_k-1)(a_k+1)\bar{\eta}_k\xi_k^T$$
$$\times(A+I)-(z+1)(a_k\bar{a}_k-1)\eta_k\xi_k^T(A+I)]$$
$$=(C-C_a^{k-1}(z)C_m^{k-1})$$
$$+\frac{(a_k\bar{a}_k-1)}{(z\bar{a}_k-1)(\bar{a}_k+1)(a_k+1)}C_a^{k-1}(z)\eta_k\xi_k^T$$
$$\times[(z+1)(\bar{a}_k+1)(a_kI-A)+(z\bar{a}_k-1)(a_k+1)$$
$$\times(A+I)-(z+1)(a_k\bar{a}_k-1)(A+I)]$$
$$=(C-C_a^{k-1}(z)C_m^{k-1})+\frac{a_k\bar{a}_k-1}{z\bar{a}_k-1}C_a^{k-1}(z)\bar{\eta}_k\xi_k^T(zI-A)$$

In the above derivation we used the fact that $\eta_k^T C_m^{k-1}=\xi_k^T(a_kI-A)$. Now by definition of $E^k(z)$ we have

$$E^k(z)=(C-C_a^k(z)C_m^k)(zI-A)^{-1}AK_f$$
$$=(C-C_a^{k-1}(z)C_m^{k-1})(zI-A)^{-1}AK_f$$
$$+\frac{a_k\bar{a}_k-1}{z\bar{a}_k-1}C_a^{k-1}(z)\eta_k\xi_k^T AK_f$$
$$=E^{k-1}(z)+\frac{a_k\bar{a}_k-1}{z\bar{a}_k-1}C_a^{k-1}(z)\eta_k\xi_k^T AK_f$$

Applying the above recursive formula and noting $E^0(z)=0$ we obtain the result (33).

APPENDIX C. PROOF OF THEOREM 3.2

From Theorem 3.1 we can see that it suffices to show that

$$u^T E(z)=0 \quad \forall u \in W^\perp$$

First, we would like to show by induction that the vector η_k as defined in (26) lies in the subspace W for $k=1,2,\ldots,l$.

For $k=1$, it is obvious since $\eta_1=w_1$ by definition. Suppose that it is true for $k=l-1$, i.e. $\eta_k\in W$ for $k=1,2,\ldots,l-1$. We need to show that it is also true for $k=l$. By Definition 2.2 we have that

$$w_l^T C(a_lI-A)^{-1}B=w_l^T G(a_l)=0 \tag{C.1}$$

From Lemma 2.1 we know that

$$G(a_l)=C_a^{l-1}(a_l)G_m^{l-1}(a_l), \tag{C.2}$$

where $C_a^{l-1}(z)$ is as defined in Lemma 3.1. It follows from (C.1) and (C.2) that

$$w_l^T C_a^{l-1}(a_l)G_m^{l-1}(a_l)=0 \tag{C.3}$$

By definition of η_l (see (26)),

$$\eta_l^T C_m^{l-1}(a_lI-A)^{-1}B=\eta_l^T G_m^{l-1}(a_l)=0 \tag{C.4}$$

Since a_l is a distinct zero, the left nullspace of $G_m^{l-1}(a_l)$ is one-dimensional. Thus, (C.3) and (C.4) imply that there exists a constant c_l such that

$$\eta_l^T=c_l w_l^T C_a^{l-1}(a_l)$$

Hence, we have

$$\eta_l=c_l C_a^{l-1}(a_l)^T w_l,$$
$$=c_l\left\{\prod_{k=1}^{l-1}\left[I-\left(\frac{a_k\bar{a}_k-1}{a_k+1}\right)\left(\frac{a_l+1}{a_l\bar{a}_k-1}\right)\eta_k\eta_k^H\right]\right\}w_l \tag{C.5}$$

After expanding the the right hand side of (C.5), one can see that η_l lies in the span of $\{\eta_1,\eta_2,\ldots,\eta_{l-1},w_l\}$ over C^m. Since $\eta_1,\eta_2,\ldots,\eta_{l-1}$ are in W by assumption, we can conclude that $\eta_l\in W$.

Now we have established the fact that vectors $\eta_1,\eta_2,\ldots,\eta_l$ belong to the subspace W. Hence, if a vector is orthogonal to W, it must be orthogonal to $\eta_1,\eta_2,\ldots,\eta_l$.

From the above arguments, one can conclude that $u\in W^\perp$ implies that u is orthogonal to $\eta_1,\eta_2,\ldots,\eta_l$. This fact together with Lemma 3.1 implies that

$$u^T E(z)=0, \quad \forall u\in W^\perp,$$

which completes the proof.

Automatica, Vol 29, No 2, pp 365–372, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0 00

Direct Control Design in Sampled-data Uncertain Systems*

ODED YANIV† and YOSSI CHAIT‡

A new direction for Z-domain design of single input/output, sampled-data uncertain systems is developed within the setting of Quantitative Feedback Theory.

Key Words—Control system synthesis, robust control; sampled-data systems

Abstract—This paper introduces a new direction for design of single input/output, sampled-data uncertain systems within the setting of Quantitative Feedback Theory (QFT). The control system consists of a continuous-time uncertain plant, a discrete-time controller connected via a sample-and-hold device and a discrete-time prefilter for reference tracking. The class of problems considered here includes robust stability, robust gain and phase margins, robust discrete-time tracking and robust continuous-time tracking. The new direction involves a QFT technique where control design is performed directly in the Z-domain. It is shown that QFT bounds can be computed in the Z-domain from a set of quadratic inequalities. A numerical example illustrates the salient features of the developed technique.

INTRODUCTION

IN THIS PAPER we consider a class of sampled-data control systems (Fig. 1) where the analog plant belongs to an uncertain set $\mathcal{P}$. The uncertainties can be both parametric and non-parametric, however, we assume that the set can be reasonably approximated by a finite number of plants

$$\mathcal{P} \approx \{P_1(s), P_2(s), \ldots, P_n(s)\}, \quad n < \infty.$$

The sample-and-hold device is assumed to be zero-order. Both the prefilter, $F(z)$, and the controller, $G(z)$, are implemented in discrete-time.

In general, design of a sampled-data control system can be performed using either a discrete-time or a continuous-time approach. The discrete-time approach employs extensions of various continuous-time control design techniques to the discrete-time setting. Naturally, control design objectives can be considered only in terms of discrete-time transfer functions. The continuous-time approach can consider continuous-time objectives (e.g. intersample behavior). Several directions have been suggested in order to make direct continuous-time design possible for this time-varying system. One direction employs frequency domain bounds on the time varying functions in Fig. 1 via conic sectors (Thompson *et al.*, 1983, 1986). A second direction is based on the following well-known approximation: the frequency response of a discrete-time signal is "equal", up the Nyquist frequency, to the frequency response of the continuous-time signal divided by the sampling time (e.g. Morari and Zafiriou, 1989). A third direction, referred to as PCT (pseudo-continuous-time), applies Padè approximation to replace the sampler and zero-order-hold by a rational continuous-time transfer function (Houpis and Lamont, 1985). A recent direction focuses on computation of induced norms of a sampled-data system (Leung *et al.*, 1989; Kabamb and Hara, 1990; Hara and Kabamba, 1990; Chen and Francis, 1991).

This paper develops a new direction for design of single input/output, sampled-data uncertain systems within the setting of traditional QFT. Recent extension of the continuous-time QFT technique to sampled-data systems were based on mapping of the QFT problem into the W-domain using bilinear transformation (Sidi, 1977; Horowitz and Liao, 1986; Tsai and Wang, 1987). They were tested using several interesting numerical examples (e.g. Schneider (1986) and Hamilton *et al.* (1989)). The technique developed in this paper is based on a direct Z-domain procedure, which resolves the warping problem due to the bilinear transformation. In

* Received 5 September 1990; revised 25 March 1991; revised 9 June 1992; received in final form 30 June 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor P. Dorato under the direction of Editor H. Kwakernaak.

† Electrical Engineering-Systems Department, Tel Aviv University, Tel Aviv 69978, Israel.
‡ Mechanical Engineering Department, University of Massachusetts, Amherst, MA 01003, U.S.A.

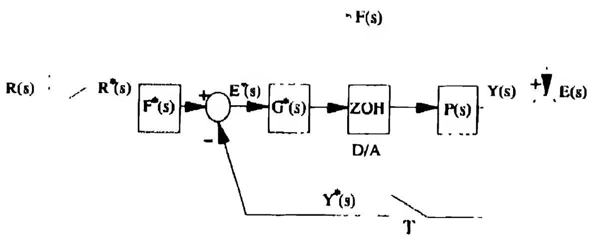


FIG. 1. A two degree-of-freedom sampled-data control system.

addition to the robust stability and robust discrete-time tracking problems investigated in Sidi (1977), Horowitz and Liao (1986) and Tsai and Wang (1987), we also consider the robust gain and phase margins and the continuous-time tracking problems.

This paper is organized as follows. A short discussion on the notation is presented. It is followed by a statement of the specific control design problems. Next, we develop the QFT design technique in the Z-domain. Finally, a numerical example is presented along with relevant discussions on: QFT bounds at different sampling times, correlation between QFT bounds for discrete-time and continuous-time tracking problem.

NOTATION

In Fig. 1, T denotes the sampling time and $\omega_s = 2\pi/T$ denotes the sampling frequency. The output of the A/D converter, modeled mathematically as an impulse modulator, fed by the signal $x(t)$ is a string of impulses

$$x^*(t) = \sum_{k=0}^{\infty} x(kT)\,\delta(t-kT).$$

The Laplace transform of this sequence is

$$X^*(s) = \sum_{k=0}^{\infty} x(kT)e^{-skT}, \qquad (1)$$

while the periodicity of $X^*(s)$ gives

$$X^*(s) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X(s - jk\omega_s).$$

Using the transformation $z = e^{sT}$, $X^*(s)$ is the Z-transform of $x(t)$, $X(z) = Z\{(x(t)\}$. The D/A converter is modeled as a zero-order hold, which constructs a piecewise continuous-time signal from discrete-time signal and has the transfer function

$$\mathrm{ZOH}(s) = \frac{1 - e^{-sT}}{s}.$$

For simplicity, we use the notation

$$P_{\mathrm{ZOH}}(s) \triangleq P(s)\mathrm{ZOH}(s).$$

Using equation (1) and $z = e^{sT}$, we have $P_{\mathrm{ZOH}}(z) = Z\mathscr{L}^{-1}\{\mathrm{ZOH}(s)P(s)\}$ for each $P(s) \in \mathscr{P}$, $R^*(s) = Z\mathscr{L}^{-1}\{R(s)\}$, $Y^*(s) = Z\mathscr{L}^{-1}\{Y(s)\}$ and $E^*(s) = E(z)$ ($\mathscr{L}^{-1}$ denotes inverse Laplace transform).

STATEMENT OF THE PROBLEM

Consider the sampled-data system shown in Fig. 1. The controller, $G(z)$, assumed to be implemented directly in the Z-domain as a different equation. The discretized version of the analog prefilter $F(s)$ is derived using the emulation $F(z) = Z\mathscr{L}^{-1}\{F(s)\mathrm{ZOH}(s)\}$ (the results in this paper also hold when $F(z)$ is obtained using other emulation methods).

The following closed-loop objectives are considered in this paper: for each $P(s) \in \mathscr{P}$ and $\omega \in [0, \omega_s/2]$ $(z = e^{j\omega T})$.

- Robust stability.
- Robust gain and phase margins:

$$|1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)| \geq \mu.$$

- Robust discrete-time tracking:

$$\frac{E^*(j\omega)}{|R^*(j\omega)|} \leq \delta_1(\omega).$$

- Robust continuous-time tracking:

$$\frac{E(j\omega)}{R(j\omega)} \leq \delta_2(\omega).$$

The control design problem is: given the set $\mathscr{P}$, and the prefilters, $F(s)$ and $F(z)$, find an appropriate discrete controller, $G(z)$, to meet some or all of the above objectives.

Remark 1. Robust stability implies that the closed-loop system is stable for each $P(s) \in \mathscr{P}$. The inequality $|1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)| \geq \mu$ is related to robust phase and gain margins as follows. Let $1 + P_{\mathrm{ZOH}}(z)G(z)$ denote a fixed characteristic polynomial of a stable discrete-time system (that is not conditionally stable). Introduce the polar form: $L^*(j\omega) = P^*_{\mathrm{ZOH}}(j\omega) = G^*(j\omega)le^{j\lambda}$ and $z = e^{j\omega T}$, $\omega \in [0, \omega_s/2]$. The gain margin is defined as $\mathrm{GM} = 1/l$ at $\lambda = -180^\circ$, hence $\mathrm{GM} = (1-\mu)^{-1}$. The phase margin of the system is defined as $\mathrm{PM} = 180^\circ + \lambda$ where λ is the phase of $L^*(j\omega)$ at $|L^*(j\omega)| = 1$, hence $\mathrm{PM} = 180^\circ + 2\cos^{-1}(\mu/2)$. If this objective is guaranteed for each $P(s) \in \mathscr{P}$, then the margins are said to be robust. Similar relations for conditionally stable systems can also be derived. For a basic discussion on gain and phase margins for fixed (i.e. without uncertainty) discrete-time feedback systems see Franklin *et al.* (1990). For a detailed presentation of Nyquist and Nichols stability analysis for fixed plants see Dorf (1989).

THE DESIGN TECHNIQUE

Robust stability

The notion of robust stability amounts to checking stability using one randomly chosen nominal loop, and then demonstrating closed-loop stability of the uncertain system by some argument involving the connected nature of $\mathcal{P}$. On a Nichols chart, this is done as follows; at each point, s, on the Nyquist contour, the responses of $L^*(j\omega) = P^*_{ZOH}(j\omega)G^*(j\omega)$, $P(j\omega) \in \mathcal{P}$, fill in a neighborhood of the nominal response $L^*_0(j\omega) = P^*_{ZOH0}(j\omega)G^*(j\omega)$, $P_0(j\omega) \in \mathcal{P}$. This neighborhood is called a *template*, which is assumed to be connected and simply connected. As s traverses the Nyquist contour, the union of these templates becomes a connected region, which we shall call the Nichols envelope. If each $P(j\omega) \in \mathcal{P}$ has the same number of unstable poles, we can invoke the following Nichols robust stability criterion of Cohen *et al.* (1992). The closed-loop system is said to be robustly stable if and only if the nominal system is stable and

$$|1 + P^*_{ZOH}(j\omega)G^*(j\omega)| > 0,$$
$$\omega \in [0, \omega_s/2], \quad \text{for each } p(j\omega) \in \mathcal{P}.$$

If the set $\mathcal{P}$ is approximated by a finite number of plants, then it is possible that some plant, not retained after the approximation, violates the above condition. To minimize such a possibility we replace the above by the more conservative constraint

$$|1 + P^*_{ZOH}(j\omega)G^*(j\omega)| > \mu > 1,$$
$$\omega \in [0, \omega_s/2], \quad \text{for each } P(j\omega) \in \mathcal{P}.$$

In essence, rather than avoiding a region defined by a single point, we consider a region defined by the classical constant magnitude curves. This condition is also used to define stability margins as follows.

Robust gain and phase margins

Traditionally, QFT design involves the mapping of the robust margins objective into certain bounds on a nominal loop transmission $L^*_0(s)$, followed by loop shaping. These bounds divide the complex-plane into two domains where $L^*_0(s)$ should lie inside one. We now show that computation of these bounds in the Z-domain can be performed in an elegant way as follows. Note that Sidi (1977), Horowitz and Liao (1986) and Tsai and Wang (1987) employed W-domain bounds. Consider the inequality

$$|1 + P^*_{ZOH}(j\omega)G^*(j\omega)| \geq \mu,$$
$$\omega \in [0, \omega_s/2], \quad \text{for each } P(j\omega) \in \mathcal{P},$$

at a fixed frequency. Convert the functions into their polar forms

$$G^*(j\omega) = g e^{j\gamma} \quad \text{and} \quad P^*_{ZOH}(j\omega) = p e^{j\theta},$$

then plug into the above inequality

$$|1 + gpe^{j(\gamma+\theta)}| \geq \mu, \quad \text{for each } P(j\omega) \in \mathcal{P},$$

finally square both sides and rearrange in terms of $g = |G^*(j\omega)|$,

$$g^2[p]^2 + 2g[p\cos(\gamma + \theta)] + [1 - \mu^2] \geq 0,$$
$$\gamma \in [-2\pi, 0], \quad \text{for each } P(j\omega) \in \mathcal{P}. \quad (2)$$

This quadratic inequality defines a map from the problem data into the QFT bounds. Numerically, this map is evaluated as follows. At each fixed z on the unit circle, a fixed controller's phase γ, and for a fixed plant from the set $\mathcal{P}$, the only unknown variable in equation (2) is the gain $g = |G^*(j\omega)|$. The two solutions of the quadratic equation, $g_{min}(\gamma)$ and $g_{max}(\gamma)$, indicate the allowable range of the controller's gain, $g(\gamma)$, necessary to satisfy the robust margins objective. typical solutions of equation (2) and their interpretation are shown in Table 1.

A QFT bound on $G^*(j\omega)$ is obtained by varying γ in the range $[-2\pi, 0]$ while keeping the frequency fixed. Similar bounds are computed for each uncertain plant yielding a set of bounds. The union of this set of bounds defines a single bound on $G^*(j\omega)$ at that frequency. This step is repeated at other frequencies yielding a set of bounds on $G^*(j\omega)$. Graphically speaking, these bounds split the Nichols chart into two regions, where $G^*(j\omega)$ must lie within one.

The final step in a QFT design involves loop shaping of $G^*(j\omega)$ to satisfy its bounds. In practice, it is recommended that one actually perform loop shaping using a nominal loop transmission, $L^*_0(j\omega)$. The bounds for $L^*_0(j\omega)$ are simply equal to the bounds for $G^*(j\omega)$ multiplied by an arbitrarily chosen nominal plant $P^*_{ZOH}(j\omega)$ from the set $\mathcal{P}$. If an $L^*_0(j\omega)$ is found such that we have nominal closed-loop stability and $L^*_0(j\omega)$ lies within its bounds, then the uncertain closed-loop system is said to have robust gain and phase margins corresponding to the value of μ.

TABLE 1. POSSIBLE SOLUTIONS OF EQUATION (2) AND THEIR INTERPRETATION ($\mu > 1$)

g_{min}	g_{max}	Conditions on g
Complex	Complex	$g > 0$
Real, ≤ 0	Real, ≤ 0	$g > 0$
Real, ≤ 0	Real, > 0	$g \geq g_{max}$
Real, > 0	Real, > 0	$g \leq g_{min}$, $g \geq g_{max}$

The discrete-time tracking objective

Using basic block diagram algebra, the discrete-time tracking objective for the system shown in Fig. 1 is given by

$$\left|\frac{E^*(j\omega)}{R^*(j\omega)}\right| = \left|F^*(j\omega) - \frac{Y^*(j\omega)}{R^*(j\omega)}\right|$$

$$= \left|F^*(j\omega) - F^*(j\omega)\frac{P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)}{1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)}\right|$$

$$\leq \delta_1(\omega),$$

which can be simplified to

$$|1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)| \geq \frac{|F^*(j\omega)|}{\delta_1(\omega)},$$

$$\omega \in [0, \omega_s/2], \quad \text{for each } p(j\omega) \in \mathcal{P}. \quad (3)$$

But this inequality has the exact form of the inequality arrived at in the robust margins objective. Therefore, computation of QFT bounds for this objective is done in the same manner as described above.

In general, a successful discrete-time tracking design may still result in a continuous-time response that exhibits intersample ripple. For this reason we now consider the continuous-time tracking objective.

The continuous-time tracking objective

For the sampled data system in Fig. 1 we have

$$Y(s) = P_{\mathrm{ZOH}}(s)G^*(s)(F^*(s)R^*(s) - Y^*(s)).$$

As expected, because the system is time-varying, it is not possible to derive a standard transfer function from $R(s)$ to $Y(s)$. Several schemes, described in the Introduction, can be used to get around this difficulty. In this paper, we propose a new scheme which does not involve approximations or conservative bounds. After several algebraic steps, the continuous-time tracking error in Fig. 1 can be written as

$$\left|\frac{E(j\omega)}{R(j\omega)}\right| = \left|F(j\omega) - \frac{Y(j\omega)}{R(j\omega)}\right|$$

$$= \left|\frac{F(j\omega)}{1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)} - \frac{F(j\omega)G^*(j\omega)}{1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)}\right.$$

$$\left.\times\left(\frac{F^*(j\omega)}{F(j\omega)}\frac{R^*(j\omega)}{R(j\omega)}P_{\mathrm{ZOH}}(j\omega) - P^*_{\mathrm{ZOH}}(j\omega)\right)\right|.$$

The above can be simplified to

$$\left|\frac{1 - G^*(j\omega)\left(\dfrac{F^*(j\omega)R^*(j\omega)}{F(j\omega)R(j\omega)}P_{\mathrm{ZOH}}(j\omega) - P^*_{\mathrm{ZOH}}(j\omega)\right)}{1 + P^*_{\mathrm{ZOH}}(j\omega)G^*(j\omega)}\right|$$

$$\leq \delta_3(\omega), \quad \omega \in [0, \omega_s/2], \quad \text{for each } P(j\omega) \in \mathcal{P}$$

where

$$\delta_3(\omega) = \frac{\delta_2(\omega)}{|F(j\omega)|}.$$

This inequality can also be transformed into a quadratic inequality in order to map the objective into its corresponding QFT bounds. Again, let us introduce polar forms, $G^*(j\omega) = ge^{j\gamma}$ and $P^*_{\mathrm{ZOH}}(j\omega) = pe^{j\theta}$, and define

$$\frac{F^*(j\omega)}{F(j\omega)}\frac{R^*(j\omega)}{R(j\omega)}P_{\mathrm{ZOH}}(j\omega) - P^*_{\mathrm{ZOH}}(j\omega) = ae^{j\alpha}.$$

Note that the particular class of reference inputs must be known in order to obtain QFT bounds. Now plug these into the above equation.

$$\left|\frac{1 - gae^{j(\gamma+\alpha)}}{1 + pge^{j(\gamma+\theta)}}\right| \leq \delta_3(\omega).$$

Next, square both sides and rearrange in terms of the controller's magnitude g $|G^*(j\nu)|$

$$g^2\left(p^2 - \frac{a^2}{\delta_3^2(\omega)}\right) + 2g\left(p\cos(\gamma - \theta) + \frac{a\cos(\gamma + \alpha)}{\delta_3^2(\omega)}\right)$$

$$+ \left(1 - \frac{1}{\delta_3^2(\omega)}\right) \geq 0, \quad \omega \in [0, \omega_s/2],$$

$$\gamma \in [-2\pi, 0], \quad \text{for each } P(j\omega) \in \mathcal{P}. \quad (4)$$

Given such a quadratic inequality, the procedure for computing QFT bounds was described earlier. It can be seen from our approach that in order to derive a transfer function from $R(s)$ to $Y(s)$, one must specify *a priori* the class of reference inputs.

Remark 2. The term

$$\frac{F^*(j\omega)}{F(j\omega)}\frac{R^*(j\omega)}{R(j\omega)}P_{\mathrm{ZOH}}(j\omega) - P^*_{\mathrm{ZOH}}(j\omega),$$

distinguishes the continuous-time tracking objective from the discrete-time tracking objective. In the limit, as $T \to 0$, equation (4) reduces to equation (3).

Simultaneous objectives

If the feedback problem consists of simultaneous objectives, such as robust margins and robust tracking objectives, then the nominal loop transmission, $L_0^*(s)$, must be shaped so to simultaneously satisfy both bounds. Note that the QFT procedure for simultaneous objectives does not lead to a conservative design since QFT bounds are computed separately for each objective.

AN EXAMPLE

Consider the sampled-data system shown in Fig. 1 with a sampling time of $T = 0.01$ sec. The

analog uncertain plant is described by the set

$$\mathcal{P} = \left\{ \frac{k}{s(s+a)},\ k \in [1, 4],\ a \in [1, 4] \right\}$$

For computing the QFT bounds, this set was approximated by

$$\mathcal{P} \approx \left\{ \frac{k}{s(s+a)},\ k \in [1, 2, 3, 4],\ a \in [1, 2, 3, 4] \right\}$$

The robust phase and gain margins objective is

$$|1 + P^*_{ZOH}(j\omega)G^*(j\omega)| \geq \tfrac{1}{2},$$
$$\omega \in [0, \omega_s/2], \quad \text{for each } P(j\omega) \in \mathcal{P},$$

which implies PM = 30° and GM = 2. The analog prefilter is given by

$$F(s) = \frac{5}{s+5},$$

and the discretized prefilter is obtained using the emulation $F(z) = Z\mathcal{L}^{-1}\{\mathrm{ZOH}(s)F(s)\}$. The discrete-time tracking error bound is

$$\delta_1(\omega) = \frac{\omega}{10.75}$$

The robust stability and robust discrete-time tracking bounds, computed from the maps defined by equations (2) and (3) are shown in Fig. 2. For an arbitrary choice of the nominal plant

$$P_0(s) = \frac{1}{s^2 + s},$$

a synthesized nominal loop transmission $L_0^*(s) = P^*_{ZOH}(j\omega)G^*(j\omega)$ which lies within its bounds

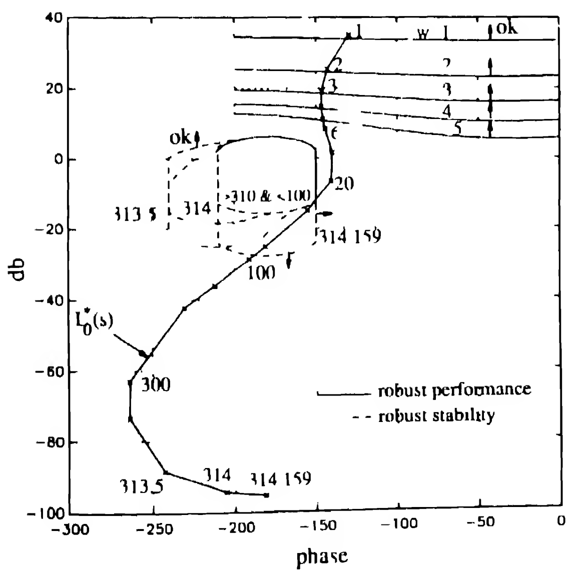


FIG. 2. Robust margins and discrete-time tracking bounds and the nominal loop transmission for $T = 0.01$ and first-order prefilter.

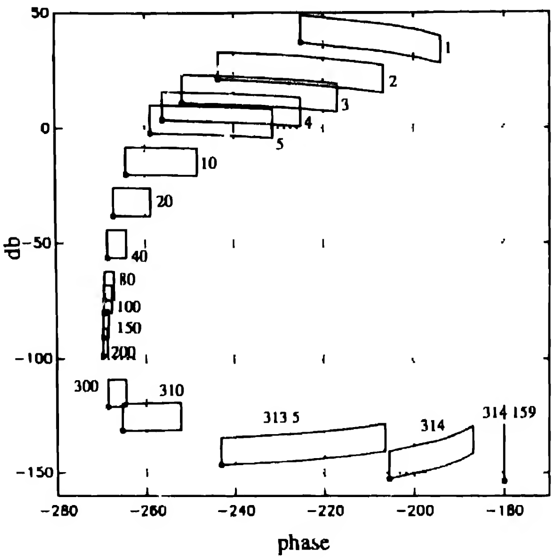


FIG. 3. Templates of $P^*(j\omega)$ for $T = 0.01$.

and achieves nominal closed-loop stability is shown in Fig. 2. The controller is given by

$$G(z) = 80\frac{z - 0.92}{z - 0.65}$$

The plant templates at various frequencies are depicted in Fig. 3.

This direct Z-domain technique offers useful insight into the tradeoffs between design objectives and controller complexity. For example, from Fig. 2 it is clear that the nominal $L_0^*(s)$ lies very close to its bounds in the range $\omega \in [1, 10]$. This implies that the tracking error $|E^*(j\omega)|$ is very close to its weight $\delta_1(\omega)$ at that frequency range. This fact is verified in Fig. 4, where it can also be observed that the objective is actually violated at $\omega = 9$, a frequency at

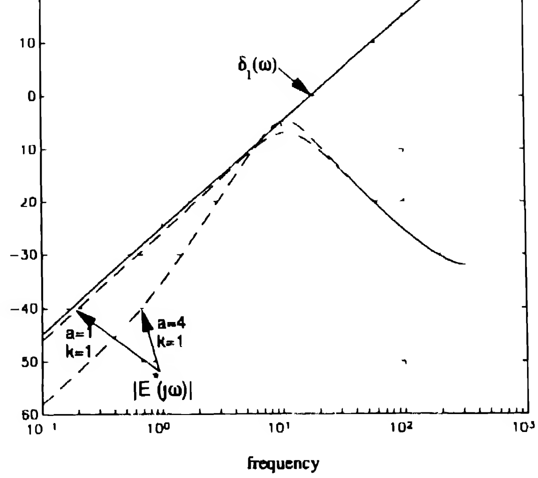


FIG. 4. Closed-loop discrete-time tracking error vs weight.

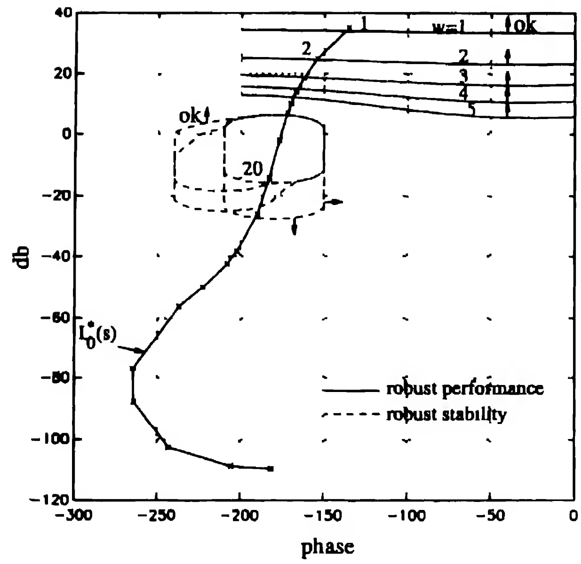


Fig 5 Figure 2 with the controller replaced by $G(z) = 80$

which bounds were not computed This frequency by frequency design procedure can reveal to the designer where changes in the weight at some frequency range can lead to a more economical design (i e lower bandwidth or reduced controller's complexity) Moreover, suppose the designer wishes to quickly test whether a controller with a certain structure can solve the problem For example, it is easy to show that the simple controller $G(z) = k$ cannot simultaneously achieve robust margins and robust discrete-time tracking to see this, we only need to plot the nominal $L_0^*(s)$ over the QFT bounds as done in Fig 5 (with $k = 80$)

Returning to analyse the control design, the discrete step responses (corresponding to all plants) are relatively "close" to each other, however, they are not as "fast" as the step response of the discrete prefilter (Fig 6) This apparent inconsistency is related to the fact that precise translation from time domain specifications into appropriate frequency domain specifications is not possible Any frequency domain based design technique is subjected to this limitation Of course, certain approximations and bounds can be used to add rigor to this translation In this example, a tracking weight based on a first-order prefilter was chosen for a second order plant, complicating the translation related inaccuracies A more reasonable prefilter is given by

$$F(s) = \frac{20\ 25}{s^2 + 3\ 15s + 20\ 25}$$

The emulated discrete-time prefilter is given by $F(z) = Z\mathcal{L}^{-1}\{\mathrm{ZOH}(s)F(s)\}$ The corresponding QFT bounds for the robust margins and robust discrete-time tracking problems along with the designed nominal $L_0^*(s)$ are shown in Fig 7 The performance bounds coresponding to the new prefilter are tougher around $\omega = 3$ requiring larger loop gain The controller is given by

$$G(z) = 110\frac{z - 0\ 9}{z - 0\ 55}$$

As expected, this choice of prefilter results in acceptable discrete step responses (see Fig 8)

The continuous-time tracking bounds for a step input with the weight $\delta_2(\omega) = \delta_1(\omega)$ match closely those obtained for the discrete-time tracking objective (Fig 9) This resemblance indicates that, in this particular example, even with a controller designed based on the discrete-time tracking objective, the system's continuous-time step response will follow closely the step response of the prefilter $F(s)$ with

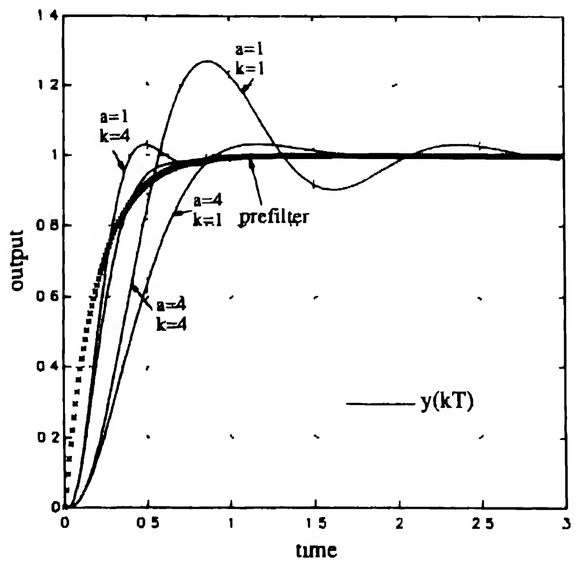


Fig 6 Closed loop time responses with first-order prefilter

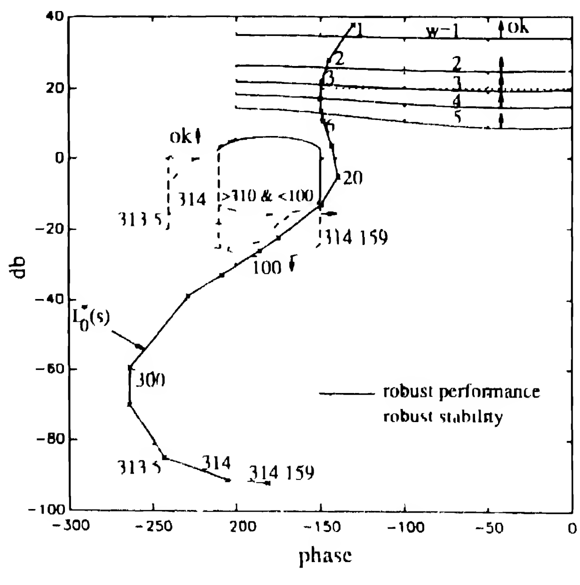


Fig 7 Robust margins and discrete-time tracking bounds and the nominal loop transmission for $T = 0\ 01$ and second order prefilter

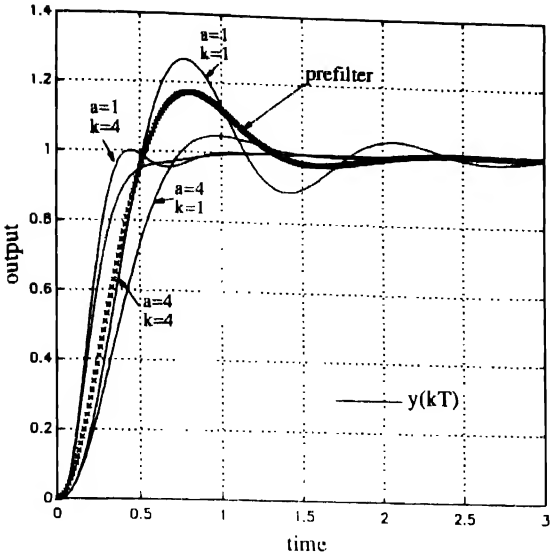


FIG. 8. Closed-loop time responses with second-order prefilter.

negligible intersample ripple. Naturally, an improperly designed controller which leads to a pole/zero cancellation close to $z = 1$ in $G(z)P_{ZOH}(z)$, may result in an intersample ripple (e.g. Morari and Zafiriou, 1989).

Finally, we have compared the QFT bounds for discrete-time tracking at different sampling times of $T = 0.01$ and $T = 0.5$. Because the frequency range $\omega < 5$ is far below either sampling frequencies, the templates of $P^*(s)$ at either sampling times have similar shapes. This implies that we should expect here similar bounds at these have similar shapes. This implies that we should expect here similar bounds at these two sampling times, as verified in Fig. 10. Nevertheless, similar bounds at different sampling times do not necessarily imply that the

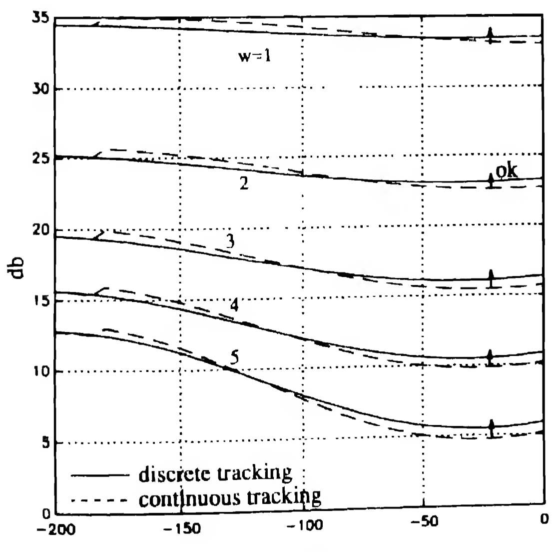


FIG. 9. A comparison between discrete-time and continuous-time tracking bounds for $T = 0.01$.

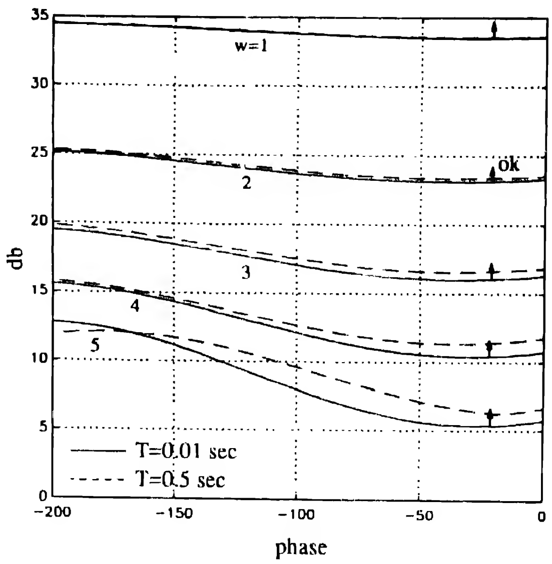


FIG. 10. A comparison between discrete-time tracking bounds for $T = 0.01$ and $T = 0.05$.

same controller can be used for both cases. Indeed, the system with $T = 0.5$ will require more phase lead in the controller compared with the $T = 0.01$ system due to the added lag in $P^*_{ZOH}(s)$.

CONCLUSIONS

A new control design technique for sampled-data uncertain systems was developed in the spirit of continuous-time, single input/output QFT. Using this technique, control design is performed directly in the Z-domain for the robust stability, robust gain and phase margins, robust discrete-time tracking and robust continuous-time tracking problems. It was shown that Z-domain QFT bounds can be easily computed from a set of quadratic inequalities.

Acknowledgements—This research was supported by a United States–Israel Binational Science Foundation Grant No. 86-00034 at Tel Aviv University, Israel, and by a National Science Foundation Grant No. MSS-8920628 at the University of Massachusetts, U.S.A.

References

Chen, T. and B. A. Francis (1990). On the L_2-induced norm of a sampled-data system. *Systems and Control Letters*, **15**, 211–219.

Cohen, N., Y. Chait, O. Yaniv and C. Borghesani (1992). Stability analysis using Nichols charts. *1st QFT Symposium*, (to be presented).

Dorf, C. R. (1989). *Modern Control Systems*. Addison-Wesley, New York.

Franklin, G. F., J. D. Powell and M. L. Workman (1990). *Digital Control of Dynamic Systems*. Addison-Wesley, New York.

Hamilton, S., I. M. Horowitz and C. H. Houpis (1989). QFT digital controller for an unmanned research vehicle (URV). *Proc. ACC*, pp. 2441–2452.

Hara, S. and P. T. Kabamba (1990). Worst case analysis and design of sampled-data control system. *Proc. CDC*, pp. 202–203.

Horowitz, I. M. and Y. K. Liao (1986). Quantitative

feedback design for sampled-data systems. *Int. J. Control*, **44**, 665–675.

Houpis, C. H. and G. B. Lamont (1985). *Digital Control Systems*. McGraw Hill, New York.

Kabamba, P. T. and S. Hara (1990). On computing the induced norm of a sampled data system. *Proc. ACC*, pp. 319–320.

Leung, G. M. H., T. P. Perry and B. A. Francis (1989). Performance analysis of sampled-data control systems. Systems Control Group Report No. 8916, University of Toronto, Canada.

Morari, M. and E. Zafiriou (1989). *Robust Process Control*, Prentice-Hall, Englewood Cliffs, NJ.

Schneider, D. L. (1986). QFT ditital flight control design as applied to the AFTIF/F-16. MS thesis, School of Engineering, Air Force Institute of Technology, Wright-Patterson, AFB, AFIT/GE/ENG(86D-4).

Sidi, M. (1977). Synthesis of sampled feedback systems for prescribed time domain tolerances. *Int. J. Control*, **26**, 445–461.

Thompson, P. M., G. Stein and M. Athans (1983). Conic sectors for sampled-data feedback systems. *System and Control Letters*, **3**, 77–82.

Thompson, P. M., R. L. Dailey and J.C. Doyle (1986). New conic sectors for sampled-data feedback systems. *System and Control Letters*, **7**, 395–404.

Tsai, T-E. and T.-S. Wang (1987). Optimal design of digital control systems with large plant uncertainty. *Int. J. Control*, **46**, 201–214.

Automatica, Vol 29, No 2 pp 373 385 1993
Printed in Great Britain

0005 1098/93 $6 00 + 0 00

The Generalized H_2 Control Problem*

MARIO A. ROTEA†

The problem of finding a controller such that the closed loop gain from $L_2[0, \infty)$ *to* $L_\infty[0, \infty)$ *is below a specified level is solved by converting it into a convex optimization problem over a finite-dimensional space.*

Key Words—Control system synthesis, convex programming, disturbance rejection, linear optimal control

Abstract—In this paper we consider the problem of finding an internally stabilizing controller such that the controlled or regulated signals have a guaranteed maximum peak value in response to arbitrary (but bounded) energy exogenous inputs More specifically, we give a complete solution to the problem of finding a stabilizing controller such that the closed loop gain from $L_2[0, \infty)$ to $L_\infty[0, \infty)$ is below a specified level We consider both state-feedback and output feedback problems. In the state-feedback case it is shown that if this synthesis problem is solvable, then a solution can be chosen to be a constant state-feedback gain Necessary and sufficient conditions for the existence of solutions as well as a formula for a state-feedback gain that solves this control problem are obtained in terms of a finite dimensional convex feasibility program. After showing the separation properties of this synthesis problem, the output feedback case is reduced to a state-feedback problem It is shown that, in the output feedback case, generalized H_2 controllers can be chosen to be observer based controllers The theory is demonstrated with a numerical example

1. INTRODUCTION

CONSIDER THE finite-dimensional linear time-invariant feedback system depicted in Fig. 1, where $\mathcal{G}$ denotes the plant and $\mathcal{C}$ the controller. The signal w denotes the exogenous input vector, while z denotes the controlled output vector. The signals u and y denote the control input vector and the measured output vector, respectively. Let $\mathcal{T}_{zw}$ denote the closed loop map from the exogeneous input w to the controlled output z.

Many important control problems can be shown to be equivalent to the problem of finding a dynamic controller $\mathcal{C}$ such that the feedback system is internally stable, and the closed loop map $\mathcal{T}_{zw}$ is small in a suitably defined sense. Theories such as H_2, H_∞, and L_1 optimal control are some of the concrete and well-known examples of this point of view. In these theories, the goal is to find a controller that minimizes, or maintains below a desirable level, a given norm on $\mathcal{T}_{zw}$. Such a formulation of the synthesis problem not only captures many practical disturbance rejection/tracking specifications, but also some robust stability specifications.

Recently Wilson (1989) has introduced a number of interesting "system gains". See also Corless *et al.* (1989). Suppose that the feedback system in Fig. 1 is internally stable, and let T_{zw} denote the closed loop transfer matrix from w to z. From the results of Wilson, it follows that, if T_{zw} is strictly proper, $\mathcal{T}_{zw}$ is a bounded operator from $L_2[0, \infty)$ to $L_\infty[0, \infty)$, and its induced norm is given by

$$\|\mathcal{T}_{zw}\| = \sqrt{f\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} T_{zw}(j\omega)T_{zw}^*(j\omega)\,\mathrm{d}\omega\right)}. \quad (1)$$

The function $f(.)$ is either the maximum eigenvalue or the maximum diagonal entry, depending on the "spatial" norm used on the controlled signal $z(t)$ (for a precise definition of spatial norms see Section 2). Note that when z is a scalar signal, (1) reduces to the familiar H_2 norm of T_{zw}. On the other hand, if z is a vector-valued signal this induced norm is no longer the standard H_2 norm. Furthermore, due to the presence of the maximum eigenvalue or the maximum diagonal entry, $\|\cdot\|$ is not differentiable. In principle, this lack of differentiability could complicate the optimal control problem resulting from minimizing $\|\mathcal{T}_{zw}\|$ over all stabilizing compensators. On the other hand, the lack of this desirable property will not affect a sub-optimal approach.

Let γ denote a positive constant. The synthesis problem considered in this paper is the following.

* Received 15 February 1991; revised 17 September 1991; revised 7 May 1992; received in final form 5 June 1992. The original version of this paper was presented at the 1st IFAC Symposium on Design Methods of Control Systems which was held in Zurich, Switzerland during September 1991. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K.

† School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907, U.S.A

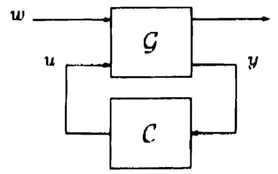


Fig. 1. The synthesis framework.

Generalized H_2 control problem. "Find (if possible) an admissible controller $\mathscr{C}$ such that the closed loop system is internally stable and the induced norm (1) (the gain from L_2 to L_∞) satisfies $\|T_{zw}\| < \gamma$."

The importance of this synthesis should be clear. It guarantees that whenever the exogenous input w has bounded energy, the controlled output z has a specified maximum peak value. As in the standard H_2 control problem, the performance measure (1) is a function of the integral of the "square" of the closed loop transfer matrix. This motivates us to call the above synthesis problem the generalized H_2 control problem.

Previous research on this problem includes the work of Wilson (1990), Grimble (1990), and Zhu and Skelton (1991). Wilson (1990) showed that the stationary Kalman filter is optimal when filtering performance is measured by the system gain (1). In this case, T_{zw} should be interpreted as the map from noise to estimation error. The "control" interpretation of this result is that the Kalman filter gain is optimal with respect to the cost (1), when "full output injection" to the state equation is available. Grimble (1990) established a connection between the standard LQG problem and a "dual" version of the generalized H_2 control problem. In our context, the results of Grimble show that an LQG controller is optimal for the generalized H_2 control problem for a plant with "larger weighting" in the regulated variables. Grimble considered only the case in which the maximum eigenvalue is used in the definition of the induced norm. Zhu and Skelton (1991) also approached the generalized H_2 problem via the solution to an LQG problem. Based on necessary conditions for optimality, they proposed an iterative scheme for finding "LQG weights" so that a solution to the LQG problem also solves the generalized H_2 problem. Their approach assumed that optimal controllers exist and have order no bigger than that of the generalized plant $\mathscr{G}$.

When the maximum diagonal entry is used in the definition of the performance measure (1), it may be shown that the generalized H_2 control problem is equivalent to a multicriterion H_2 control problem. This is because in this case the performance requirement is to keep the H_2 norm of every single row of T_{zw} below γ. This multiobjective H_2 control problem has been considered by Boyd and Barratt (1990); and Khargonekar and Rotea (1991a). In these two references the resulting multicriterion problem is reduced to a search for LQG weights without assumptions on the structure of solutions. Boyd and Barratt have also provided an algorithm that unambiguously determines the feasibility of this multiobjective H_2 problem.

In this paper we give a complete solution to the generalized H_2 control problem. More specifically, we show that:

(i) When the state of $\mathscr{G}$ (or the state and the exogenous input) is available for feedback, memoryless state-feedback gains offer the best possible performance for this synthesis problem. That is, the generalized H_2 control problem with dynamic full information feedback (i.e. plant state and exogenous input are available to the controller) has a solution if and only if the generalized H_2 problem with memoryless (constant) state-feedback is solvable.

(ii) In the case of state-feedback, a necessary and sufficient condition for the solvability of the generalized H_2 problem, as well as a formula for a solution, may be obtained via a finite-dimensional convex feasibility program. This means that there are efficient numerical algorithms to solve this class of problems.

(iii) In the case of output feedback, the generalized H_2 problem is solvable if and only if a generalized H_2 state-feedback problem, for a suitably constructed auxiliary plant, can be solved. A solution to the output feedback problem, when it exists, can be chosen to be an observer-based controller. The observer gain is given by a standard Kalman filter, while the state-feedback gain is a solution of the generalized H_2 synthesis problem corresponding to the auxiliary plant. As a result, the output feedback case is no more difficult than the state-feedback problem.

In addition, we formulate the convex program corresponding to the generalized H_2 control problem with state-feedback in a convenient form for finding numerical solutions. In this paper, the resulting convex program is solved with the ellipsoid algorithm. Due the convexity properties of our approach, this algorithm cannot fail in determining whether the generalized H_2 problem is solvable. The algorithm may also be used to isolate, to any desired accuracy, a controller that nearly optimizes the generalized

H_2 performance measure. Finally, we give a numerical example that illustrates the theory.

The notation used in this paper is fairly standard. For a given matrix A, A' denotes its transpose and A^* its conjugate transpose. If $A = A^*$, i.e. A is hermitian, its maximum eigenvalue and diagonal entry are denoted by $\lambda_{\max}(A)$ and $d_{\max}(A)$, respectively. If A and B are hermitian matrices, $A \geq B$ (resp. $A > B$) denotes $A - B$ positive semidefinite (resp. definite). Given an hermitian matrix A, the function

$$f_r(A) := \begin{cases} \lambda_{\max}(A) & \text{if } r = 2 \\ d_{\max}(A) & \text{if } r = \infty, \end{cases}$$

is extensively used throughout this paper. Note that this function is nondecreasing. That is, $A \geq B$ implies $f_r(A) \geq f_r(B)$. Further, $f(.)$ is convex. Linear time-invariant systems are described by state-space models and they are denoted by "script" symbols, while the corresponding transfer matrices are denoted by italics. For instance, $\mathcal{G}$ denotes a system with transfer matrix G. The Hardy space H_2 consists of matrix-valued functions that are square integrable on the imaginary axis with analytic extension into the right half plane. The norm on this space is defined in the usual way and it is denoted by $\|\cdot\|_2$.

2. THE GENERALIZED H_2 CONTROL PROBLEM

In this section we define the generalized H_2 cost and give a time-domain characterization of the generalized H_2 performance measure that is useful for establishing some of our results. We conclude this section with a precise formulation of the generalized H_2 control problem

2.1. *Analysis*

Consider a finite-dimensional linear time-invariant system described by the state-space model:

$$\mathcal{T} := \begin{aligned} \dot{x} &= Fx + Gw,\ x(0) = 0 \\ z &= Hx + Jw. \end{aligned} \tag{2}$$

The matrices F, G, H, and J are real and of compatible dimensions. Suppose that $\mathcal{T}$ is internally stable, i.e. F has all eigenvalues in the open left half complex plane. Let T_{zw} denote the transfer matrix from w to z. Recall that $\|T_{zw}\|_2 < \infty$ if and only if $J = 0$. In this case, we may define

$$S(T_{zw}) := \frac{1}{2\pi} \int_{-\infty}^{\infty} T_{zw}(j\omega) T^*_{zw}(j\omega)\, d\omega \geq 0. \tag{3}$$

The generalized H_2 performance measure (or cost) for the linear time-invariant system $\mathcal{T}$ that will be considered in this paper is defined as follows:

$$f_r(S(T_{zw})) := \begin{cases} \lambda_{\max}(S(T_{zw})) & \text{if } r = 2 \\ d_{\max}(S(T_{zw})) & \text{if } r = \infty. \end{cases} \tag{4}$$

Note that $f_r(S(T_{zw})) \geq 0$. Also, if $f_r(S(T_{zw}))$ is finite, we must have $J = 0$ in (2). As in the definition of the H_2 norm, $f_r(S(T_{zw}))$ is defined in terms of the integral (3). This motivates the name "generalized H_2 cost".

Wilson (1989) showed that the performance measure defined in (4) gives rise to the following system gains:

$$\sup \quad \|_{\infty,r}/\|w\|_2 : w \in L_2[0, \infty),\ w \neq 0\}$$
$$= \sqrt{f_r(S(T_{zw}))}, \tag{5}$$

where

$$_{,r} := \sup\{\|z(t)\|_r : t \in [0, \infty)\}, \tag{6}$$

$$\|w\|_2 := \sqrt{\int_0^{\infty} w'(t) w(t)\, dt}. \tag{7}$$

In (6), the r-spatial norm is given by

$$\|z(t)\|_r := \begin{cases} & \text{if } r = 2 \\ \max\{|z_j(t)| : j = 1, \ldots, s\} & \text{if } r = \infty, \end{cases} \tag{8}$$

where s is the number of components of z. Note that the subscript 'r' in (4) stands for the spatial norm used on the output $z(t)$.

When $f_r(S(T_{zw}))$ is finite (i.e. in (2), $J = 0$) it can be easily computed in time domain as follows. Let L_c denote the controllability gramian of the pair (F, G). That is, L_c is the unique solution to

$$FL_c + L_cF' + GG' = 0.$$

Then

$$S(T_{zw}) = HL_cH',$$

which implies that $f_r(S(T_{zw})) = f_r(HL_cH')$. The following result provides an alternative characterization for the generalized H_2 performance measure that will be useful for establishing some of the results in this paper.

Lemma 2.1. Consider the system $\mathcal{T}$ defined in (2) and let T_{zw} denote the transfer matrix from w to z. Let $\gamma > 0$ be given. Then, $\mathcal{T}$ is internally stable and $f_r(S(T_{zw})) < \gamma$ if and only if $J = 0$ and there exists $Y = Y' > 0$ such that

$$FY + YF' + GG' < 0, \tag{9}$$

$$f_r(HYH') < \gamma. \tag{10}$$

The proof of this result is simple and it will be omitted. Details may be found in Rotea (1991). For a proof of a similar result, see Lemma 2.1 in Khargonekar and Rotea (1991b).

2.2. *Synthesis*

Consider the finite-dimensional linear time-invariant feedback system depicted in Fig. 1, where $\mathscr{G}$ is the plant (including frequency dependent weights) and $\mathscr{C}$ is the controller to be designed. In this paper, $\mathscr{G}$ and $\mathscr{C}$ are given by state-space models. The transfer matrices of the plant and the controller are denoted by G and C, respectively. We denote the closed loop transfer matrix by T_{zw}.

A controller $\mathscr{C}$ is called admissible (for the plant $\mathscr{G}$) if $\mathscr{C}$ internally stabilizes the plant $\mathscr{G}$. The set of all admissible controllers $\mathscr{C}$ for the plant $\mathscr{G}$ is denoted by $\mathscr{A}(\mathscr{G})$. Here "$\mathscr{A}$" stands for admissible. Note that $\mathscr{A}(\mathscr{G}) \neq \emptyset$ if and only if $\mathscr{G}$ is stabilizable from u and detectable from y.

Let $\gamma > 0$ be given. The generalized H_2 controller synthesis problem is defined as follows:

"Find (if possible) a controller $\mathscr{C} \in \mathscr{A}(\mathscr{G})$ such that $f_r(S(T_{zw})) < \gamma$ ($r = 2$ or ∞)",

where $S(.)$ and $f_r(.)$ are given by (3) and (4), respectively.

Note that $\mathscr{C} \in \mathscr{A}(\mathscr{G})$ solves this synthesis problem if and only if the closed loop system is internally stable, and there exists a positive number ϵ such that the following disturbance attenuation condition holds. Given any $w \in L_2[0, \infty)$, and zero initial conditions,

$$\|z\|_{*,*} \leq (\gamma - \epsilon)\, \|w\|_2.$$

This is a simple consequence of the signal interpretation of the generalized H_2 cost (cf. (5)–(8)) given by Wilson (1989).

The objective of this paper is to show that conditions for the solvability of the generalized H_2 synthesis problem, as well as the computation of a solution (when one exists) can be obtained in terms of a finite-dimensional convex feasibility program.

In Section 3 we will be interested in memoryless (i.e. static) controllers for solving the above synthesis problem. In such a case, the generalized H_2 synthesis problem is defined in the following way. Define first the set of memoryless stabilizing controllers

$$\mathscr{A}_m(\mathscr{G}) := \{\mathscr{C} \in \mathscr{A}(\mathscr{G}) : C \in R^{q \times p}\},$$

where $q = \dim(u)$ and $p = \dim(y)$ (see Fig. 1). Then, the memoryless synthesis problem is:

"Find (if possible) a controller $\mathscr{K} \in \mathscr{A}_m(\mathscr{G})$ such that $f_r(S(T_{zw})) < \gamma$ ($r = 2$ or ∞)".

In $\mathscr{A}_m(.)$, the subscript "m" stands for memoryless controllers. (In this case it is certainly possible to identify $\mathscr{C}$ with C and think of $\mathscr{A}_m(\mathscr{G})$ as a set of matrices. However, for consistency of notation, we will use $\mathscr{C}$ when we want to think of the static gain as a system, and C when we want it to be interpreted as a matrix!)

Remark 2.1. Consider the feedback system shown in Fig. 1. Given a plant $\mathscr{G}$ and an internally stabilizing controller $\mathscr{C}$, the closed loop transfer matrix T_{zw} is a function of the individual transfer matrices G and C. Thus, in the sequel we use the notation

$$T_{zw}(G, C), \quad \text{and} \quad S(G, C) := S(T_{zw}(G, C)),$$

to indicate on which plant and controller these closed loop quantities depend.

3 STATE-FEEDBACK AND FULL INFORMATION PROBLEMS

In this section we consider the case in which the plant to be controlled is given by a state-space model where either the state vector or the state and the exogenous input are measured (full information structure). We show that the generalized H_2 synthesis problem with dynamic full information feedback has a solution if and only if the generalized H_2 synthesis problem with memoryless (constant) state-feedback has a solution. This result is analogous to the corresponding results for the LQR problem (Kalman, 1960) the standard H_∞ control problem (Khargonekar *et al.*, 1988; Doyle *et al.*, 1989) and more recently for a class of mixed H_2/H_∞ problems (Khargonekar and Rotea, 1991b). This section concludes with a convex optimization approach to the constant state-feedback problem.

Consider the generalized H_2 synthesis problem defined in Section 2 for the following plants (see also Fig. 1).

(1) State-feedback plant. The plant "$\mathscr{G}$" is given by the state-space model

$$\mathscr{G}_{sf} := \begin{cases} \dot{x} = Ax + B_1 w + B_2 u \\ z = Cx + Du \\ y = x. \end{cases} \tag{11}$$

(2) Full information plant. The plant "$\mathscr{G}$" is given by the state-space model

$$\begin{aligned} \dot{x} &= Ax + B_1 w + B_2 u \\ z &= Cx + Du \\ y &= [x'w']'. \end{aligned} \tag{12}$$

All the matrices introduced in (11) and (12) are real and of compatible dimensions. As before, we let G_{sf} and G_{fi} denote the transfer matrices of (11) and (12), respectively. The subscripts "sf" and "fi" denote state-feedback and full information structure, respectively. Note that the state-feedback and full information plants both have the same dynamical equation. The only

difference between them is in the measurement equation. We assume that the pair (A, B_2) is stabilizable. This assumption is clearly necessary for the existence of an admissible controller.

Theorem 3.1. Consider the systems $\mathcal{G}_{sf}$ and $\mathcal{G}_{fi}$ defined in (11) and (12), respectively. Let $\gamma > 0$ be given and consider the sets of admissible controllers $\mathcal{A}(\mathcal{G}_{fi})$ and $\mathcal{A}_m(\mathcal{G}_{sf})$. Then, the following statements are equivalent:

(1) there exists $\mathcal{C} \in \mathcal{A}(\mathcal{G}_{fi})$ such that $f_r(S(G_{fi}, C)) < \gamma$;
(2) there exists $\mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf})$ such that $f_r(S(G_{sf}, K)) < \gamma$.

Theorem 3.1 shows that, for full information or state-feedback, the generalized H_2 cost cannot be improved upon by the use of dynamic compensation. The proof of this theorem is very similar to the proof of Theorem 4.1 in Khargonekar and Rotea (1991b). Indeed, the techniques used by Khargonekar and Rotea also work in our context, provided that the (nondecreasing) function $f_r(.)$ is used in the definition of the performance measure. A proof of Theorem 3.1 may be found in Rotea (1991).

Next, we will develop a convex programming approach for solving the state-feedback problem. More specifically, given the plant $\mathcal{G}_{sf}$ defined in (11), we give a necessary and sufficient condition for the existence of an admissible state-feedback gain K such that $f_r(S(G_{sf}, K)) < \gamma$. This condition is given in terms of a set of inequalities. We also show that this set of inequalities defines a convex subset of a finite-dimensional space of real matrices. Hence, the problem of checking whether this set of inequalities has a solution is readily solved via convex programming techniques. Furthermore, the construction of a real matrix K such that $f_r(S(G_{sf}, K)) < \gamma$ is straightforward from our necessary and sufficient condition.

With reference to the plant $\mathcal{G}_{sf}$ defined in (11), let $n = \dim(x)$ and $q = \dim(u)$. Let $(W, Y) \in R^{q\times n} \times R^{n\times n}$ be given and define

$$L(W, Y) := AY + YA' + B_2W + W'B_2' + B_1B_1'. \tag{13}$$

If Y is nonsingular, we define also

$$M(W, Y) := (CY + DW)Y^{-1}(CY + DW)'. \tag{14}$$

Theorem 3.2. Consider the system $\mathcal{G}_{sf}$ defined in (11) and let $\mathcal{A}_m(\mathcal{G}_{sf})$ denote the set of admissible memoryless controllers for this plant. Let G_{sf} denote the transfer matrix of $\mathcal{G}_{sf}$. Let $\gamma > 0$ be given.

(1) There exists $\mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf})$ such that $f_r(S(G_{sf}, K)) < \gamma$ if and only if there exists $(W, Y) \in R^{q\times n} \times R^{n\times n}$ such that the following condition is satisfied:

$$\begin{cases} Y = Y' > 0 \\ L(W, Y) < 0 \\ f_r(M(W, Y)) < \gamma. \end{cases} \tag{15}$$

Moreover, if this condition holds, the state-feedback gain $K_0 := WY^{-1}$ satisfies

$$\mathcal{K}_0 \in \mathcal{A}_m(\mathcal{G}_{sf}) \quad \text{and} \quad f_r(S(G_{sf}, K_0)) < \gamma.$$

(2) If $\operatorname{Im} B_2 \subseteq \operatorname{Im} B_1$ and (A, B_1) is controllable, then (15) may be replaced by the condition

$$\begin{aligned} &Y = Y' > 0 \\ &L(W, Y) = 0 \\ &f_r(M(W, Y)) < \gamma. \end{aligned} \tag{16}$$

The significance of Theorem 3.2 is explained by the next result.

Theorem 3.3. Let Σ denote the set of symmetric n by n matrices, and $\Omega := \{(W, Y) \in R^{q\times n} \times \Sigma : Y > 0\}$. Consider the mappings $L : R^{q\times n} \times \Sigma \to \Sigma$, and $M : \Omega \to \Sigma$ defined in (13) and (14), respectively. Then, L is an affine mapping and M is a convex mapping with respect to the cone of positive semidefinite matrices. Consequently, the function $f_r(M) : \Omega \to R$ and the sets defined by conditions (15) and (16) are convex.

This theorem clearly shows the advantages of using condition (15), or (16), to determine whether the generalized H_2 control problem is solvable. Note that finding a pair of real matrices (W, Y) that satisfies either one of these conditions, is equivalent to a convex feasibility problem. This means that there are efficient numerical algorithms to find such a point, or determine that none exists. On the other hand, the more direct approach of searching for a state-feedback gain $\mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf})$ such that $f_r(S(G_{sf}, K)) < \gamma$ gives rise to a nonlinear programming problem which need not be convex. This is simply because the set of admissible static gains $\mathcal{A}_m(\mathcal{G}_{sf})$ does not have guaranteed convexity properties.

Note that while $\mathcal{A}_m(\mathcal{G}_{sf})$ is a subset of $R^{q\times n}$, the set defined by condition (15) is contained in a space of dimension $qn + n(n-1)/2$. On the other hand, when the technical assumption in part 2 of Theorem 3.2 holds, the dimension reduces to qn for this is the dimension of the kernel associated with the linear equation $L(W, Y) = 0$ in (16). This fact will be used later in Section 5, where the numerical solution to the generalized H_2 problem is considered.

Remark 3.1. In the next section, it will be seen that the generalized H_2 synthesis problem in the case of output feedback can be reduced to a full information/state-feedback problem similar to the one considered before. More specifically, given $R = R' \geq 0$ and letting $\mathcal{G}$ denote a full information or state-feedback plant, we will need to solve the following problem:
"Find (if possible) a controller $\mathcal{C} \in \mathcal{A}(\mathcal{G})$ such that $f_r(R + S(G, C)) < \gamma$".
It is fairly easy to show that Theorem 3.1 holds true for this "more general" synthesis problem. Moreover, Theorems 3.2–3.3 are still valid provided that the following modification is made; redefine the map $M: \Omega \to \Sigma$ introduced in (14) according to

$$M(W, Y) := R + (CY + DW)Y^{-1}(CY + DW)'.$$

Thus, the approach developed in this section is directly applicable to the more general output feedback problem.

We conclude this section with the proofs of Theorems 3.2 and 3.3. The key idea for establishing Theorem 3.2 is to replace the search over $K \in \mathcal{A}_m(\mathcal{G}_{sf})$ by a search over the set of matrices defined by conditions (15) or (16). This is done by using Lemma 2.1 in combination with the change of variables $K = WY^{-1}$, where Y is a solution to the linear matrix inequality (9). A similar result was established in Khargonekar and Rotea (1991b). For the use of this change of variables in robust stabilization problems, see Bernussou *et al.* (1989) and Packard and Becker (1990).

Proof of Theorem 3.2. Let (W, Y) denote a real solution to (15), or (16). Since Y is positive definite we may define the real matrix $K_0 := WY^{-1}$. We will show that the state-feedback gain K_0 satisfies

$$\mathcal{K}_0 \in \mathcal{A}_m(\mathcal{G}_{sf}) \quad \text{and} \quad f_r(S(G_{sf}, K_0)) < \gamma.$$

Apply the control law $u = \mathcal{K}_0 x$ to $\mathcal{G}_{sf}$. The resulting closed loop system is given by

$$\begin{cases} \dot{x} = Fx + B_1 w \\ z = Hx, \end{cases} \tag{17}$$

where $F := A + B_2 K_0$ and $H := C + DK_0$. It is now easy to show, using (13) and (14), that

$$FY + YF' + B_1 B_1' = L(W, Y),$$
$$HYH' = M(W, Y).$$

Now, suppose that (W, Y) is a solution to (15). Since $Y > 0$ satisfies the conditions of Lemma 2.1, we conclude that $\mathcal{K}_0 \in \mathcal{A}_m(\mathcal{G}_{sf})$ and $f_r(S(G_{sf}, K_0)) < \gamma$.

If (W, Y) is a solution to (16), we proceed as follows. First, note that $\operatorname{Im} B_2 \subseteq \operatorname{Im} B_1$ and (A, B_1) controllable imply that, given any K, the pair $(A + B_2 K, B_1)$ is controllable. Hence, (F, B_1) is controllable. Since $Y > 0$ satisfies $L(W, Y) = 0$, we conclude that F is asymptotically stable. Note also that Y is the controllability gramian of the pair (F, B_1). This fact, together with $f_r(HYH') = f_r(M(W, Y)) < \gamma$, imply $\mathcal{K}_0 \in \mathcal{A}_m(\mathcal{G}_{sf})$ and $f_r(S(G_{sf}, K_0)) < \gamma$.

Now we establish the necessity of (15). Suppose that there exists $\mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf})$ such that $f_r(S(G_{sf}, K)) < \gamma$. With this controller the closed loop system is given by (17), where $F := A + B_2 K$ and $H := C + DK$. From Lemma 2.1 it follows that there exists $Y = Y' > 0$ such that

$$FY + YF' + B_1 B_1' < 0, \tag{18}$$

$$f_r(HYH') < \gamma. \tag{19}$$

Define $W := KY$. Then, from (18)–(19), it immediately follows that (W, Y) satisfies (15).

The necessity of (16) is proven in a similar way. Indeed, given $\mathcal{K} \in \mathcal{A}_m(G_{sf})$, we may conclude that the controllability gramian, say Y, of the pair $(A + B_2 K, B_1)$ is positive. Using an argument similar to the one above, and the condition $f_r(S(G_{sf}, K)) < \gamma$, the necessity of (16) is established.

Theorem 3.3 is similar to Theorem 4.3 in Khargonekar and Rotea (1991b). The main difference is that Khargonekar and Rotea have shown that the scalar valued function trace (M) is convex, while here we show a more general result. Namely, the convexity of the matrix-valued function M.

Proof of Theorem 3.3. The fact that the map $L(.)$ introduced in (13) is affine follows from its definition. The convexity the map M follows from Marshall and Olkin (1979). Indeed, the domain Ω, of definition of M, is clearly convex. Note also that M is the composition of two maps, the linear map $(W, Y) \to CY + DW$ and the map $(X, Y) \to XY^{-1}X'$. The convexity of this latter map is in Marshall and Olkin (1979). We may now conclude that M is convex. The convexity of the function $f_r(M)$ follows from the convexity of M and the fact that $f_r(.)$ is nondecreasing and convex. Finally, the convexity of the sets defined by (15)–(16) follows from the properties shown for their defining functions.

4. OUTPUT FEEDBACK

In this section we consider the generalized H_2 control problem defined in Section 2 for the case of output feedback. The main result of this section (Theorem 4.1) shows that the output

feedback problem reduces to a state-feedback problem. More specifically, Theorem 4.1 shows that a generalized H_2 controller, if it exists, can be chosen to be a combination of a standard Kalman filter and a state-feedback gain for the generalized H_2 synthesis problem of a suitably constructed auxiliary plant.

Suppose that the plant in Fig. 1 is given by the following state-space model:

$$\mathcal{G} := \begin{cases} \dot{x} = Ax + B_1 w + B_2 u \\ z = C_1 x + D_1 u \\ y = C_2 x + D_2 w, \end{cases} \qquad (20)$$

where all the matrices in (20) are real matrices of compatible dimensions. As before, we let G denote the transfer matrix of (20). Note that there is no feedthrough term from the exogenous signal w to the controlled output z. Similarly, there is no direct feedthrough from the control input u to the measured output y. Although it is possible to include these terms, we have chosen not to do so in order to keep the presentation as simple as possible. In addition, we will make the following assumptions.

Assumption 1. The triple (C_2, A, B_2) is stabilizable and detectable.

Assumption 2. D_2 has full row rank and $D_2[B_1' \; D_2'] = [0 \; I]$.

Assumption 3. The pair (A, B_1) has no uncontrollable modes on the imaginary axis.

Clearly Assumption 1 is necessary, otherwise the set of admissible controllers is empty. Assumption 2 is a standard assumption in filtering theory. It ensures that there is full measurement noise and that there is no correlation between process and measurement noise. Finally, in the view of this, Assumption 3 together with (C_2, A) detectable constitutes a necessary and sufficient condition to guarantee that the Kalman filter, that estimates the state of $\mathcal{G}$, is stable. Note that the full rank property of D_2 implies that the orthogonality assumption can be made with no loss of generality. Indeed, a preliminary output injection transformation, will enforce $D_2[B_1' \; D_2'] = [0 \; I]$.

It is well known that, under the above assumptions, there exists a (unique) real symmetric matrix Q such that

$$AQ + QA' - QC_2'C_2Q + B_1B_1' = 0, \qquad (21)$$

and $A - QC_2'C_2$ is stable. Furthermore, $Q \geq 0$ and if the pair (A, B_1) is controllable then $Q > 0$. In order to give the main result of this section we define the following auxiliary system:

$$\mathcal{G}_{sf}(Q) := \begin{cases} \dot{x} = Ax + QC_2'r + B_2 u \\ v = C_1 x + D_1 u \\ y = x, \end{cases} \qquad (22)$$

where the matrices A, C_1, D_1, B_2 are as in (20). Let $G_{sf}(Q)$ denote the transfer matrix from (r, u) to (v, y) in (22). Here the notation $\mathcal{G}_{sf}(Q)$ means that this auxiliary plant depends on Q and has state-feedback structure. The main result of this section is given next.

Theorem 4.1. Consider the feedback interconnection of Fig. 1, where the plant $\mathcal{G}$ is given by (20). Suppose that Assumptions 1–3 hold. Then, there exists a unique (real symmetric) matrix Q that satisfies the ARE (21) and such that $A - QC_2'C_2$ is stable. Moreover, $Q \geq 0$. Let $\gamma > 0$ be given and consider the sets of admissible controllers $\mathcal{A}(\mathcal{G})$ and $\mathcal{A}_m(\mathcal{G}_{sf}(Q))$. Then the following statements are equivalent.

(1) There exists $\mathcal{C} \in \mathcal{A}(\mathcal{G})$ such that $f_r(S(G, C)) < \gamma$.

(2) There exists $\mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf}(Q))$ such that $f_r(C_1QC_1' + S(G_{sf}(Q), K)) < \gamma$.

Moreover, given any real matrix K that satisfies Condition 2, the dynamic output feedback controller

$$\mathcal{C}_0 := \begin{array}{l} \dot{\xi} = (A - QC_2'C_2 + B_2K)\xi + QC_2'y \\ u = K\xi, \end{array} \qquad (23)$$

satisfies

$$\mathcal{C}_0 \in \mathcal{A}(\mathcal{G}) \quad \text{and} \quad S(G, C_0) = C_1QC_1' + S(G_{sf}(Q), K),$$

which implies that $f_r(S(G, C_0)) < \gamma$.

Theorem 4.1 shows that the generalized H_2 synthesis problem with output feedback has a solution if and only if the generalized H_2 synthesis problem for the state-feedback plant $\mathcal{G}_{sf}(Q)$ has a solution. Of course, this auxiliary state-feedback problem is readily solved with the tools developed in Section 3, see Remark 3.1. Moreover, if the auxiliary state-feedback problem is solvable, a generalized H_2 controller for the plant $\mathcal{G}$ can be chosen to be an observer-based controller. In this controller, the observer gain is the Kalman filter gain $-QC_2'$, while the state-feedback gain K is a solution to the generalized H_2 synthesis problem corresponding to the auxiliary plant $\mathcal{G}_{sf}(Q)$. Although Theorem 4.1 resembles the separation structure of the standard LQG problem, there is a difference. Here, the state-feedback gain K of Condition 2 could depend on the stabilizing solution Q to the filtering ARE (21).

The proper way of using these results for solving the generalized H_2 synthesis problem in the output feedback case is summarized in the following conceptual algorithm.

(i) Find the (unique) stabilizing solution Q to the ARE (21).

(ii) Using Theorem 3.2, with the modifications indicated in Remark 3.1, determine whether the state-feedback problem for the auxiliary plant $\mathscr{G}_{sf}(Q)$ is solvable. If a solution K exists, the controller (23) solves the generalized H_2 synthesis problem for the plant $\mathscr{G}$. Otherwise, the generalized H_2 synthesis problem for the plant $\mathscr{G}$ does not have a solution.

To prove Theorem 4.1 we need a few intermediate results and definitions. First, note that Assumptions 1–3 imply that the stabilizing solution Q to the ARE (21) exists and satisfies $Q \geq 0$. Equipped with this matrix Q, we may now define the following auxiliary system

$$\mathscr{H}(Q) := \begin{cases} \dot{x} = Ax + QC_2'r + B_2u \\ v = C_1x + D_1u \\ y = C_2x + r. \end{cases} \tag{24}$$

Let $H(Q)$ denote the transfer matrix from (r, u) to (v, y) in (24).

Observe that the plant $\mathscr{G}$ defined in (20), and the auxiliary system $\mathscr{H}(Q)$, share the same triple (A, B_2, C_2). Thus, $\mathscr{A}(\mathscr{G}) = \mathscr{A}(\mathscr{H}(Q))$. We will now show that the generalized H_2 control problem for the plant $\mathscr{G}$ is equivalent to the generalized H_2 control problem for the auxiliary plant $\mathscr{H}(Q)$. Theorem 4.1 will then follow by solving this auxiliary output feedback problem. This reduction of the problem is obtained from the following lemma.

Lemma 4.2. Suppose that Assumptions 1–3 hold. Let Q denote the stabilizing solution to the ARE (21), and consider the system $\mathscr{G}$ and $\mathscr{H}(Q)$ defined by (20) and (24), respectively. Then, $\mathscr{A}(\mathscr{G}) = \mathscr{A}(\mathscr{H}(Q))$. Let an admissible controller $\mathscr{C} \in \mathscr{A}(\mathscr{G})$ be given, and consider the feedback systems shown in Fig. 2. Then, $T_{zw}(G, C) \in H_2$ if and only if $T_{vr}(H(Q), C) \in H_2$. In this case,

$$S(G, C) = C_1QC_1' + S(H(Q), C). \tag{25}$$

The significance of the separation property enjoyed by the integral $S(G, C)$ (which holds for any admissible controller $\mathscr{C}$) cannot be overemphasized. Note that (25) generalizes the well-known separation property of the scalar-valued LQG cost trace $(S(G, C))$, used in frequency-domain derivations of the LQG solution, to the matrix case.

Fig. 2. The systems of Lemma 4.2

Even though Lemma 4.2 is an easy consequence of well-established techniques (see for example, Doyle (1984) and Doyle *et al.* (1989)), it appears that the separation property of the matrix $S(G, C)$ given above has not been noted before. For the sake of brevity, the proof of this lemma will be omitted. For further details, the interested reader may consult Rotea (1991).

Proof of Theorem 4.1. First note that Assumptions 1–3 imply that the stabilizing solution Q to the ARE (21) exists and satisfies $Q \geq 0$.

We will now use Lemma 4.2 to solve the generalized H_2 control problem for the plant $\mathscr{G}$. Applying the function $f_r(.)$ to both sides of (25), we may conclude that a controller $\mathscr{C}$ satisfies Condition 1 in Theorem 4.1, if and only if $\mathscr{C}$ satisfies

$$\mathscr{C} \in \mathscr{A}(\mathscr{H}(Q)) \quad \text{and} \quad f_r(C_1QC_1' + S(H(Q), C)) < \gamma. \tag{26}$$

Hence, to solve the synthesis problem for the plant $\mathscr{G}$, it suffices to solve the generalized H_2 problem (26) for the auxiliary plant $\mathscr{H}(Q)$. Next, we show that this auxiliary output feedback problem may be solved if and only if the state-feedback problem in Condition 2 of Theorem 4.1 is solvable.

Suppose that Condition 2 in Theorem 4.1 holds, we will show that the dynamic output feedback controller $\mathscr{C}_0$ defined in (23) satisfies (26), and thus Condition 1. Note that $\mathscr{C}_0$ is an "observer-based" controller for the auxiliary plant $\mathscr{H}(Q)$. Moreover, since both $A + B_2K$ and $A - QC_2'C_2$ are stable matrices, it follows that $\mathscr{C}_0 \in \mathscr{A}(\mathscr{H}(Q))$. Note also that $T_{vr}(H(Q), C_0) = T_{vr}(G_{sf}(Q), K)$, for this observer-based controller leaves the closed loop transfer matrix T_{vr} invariant. Hence, we obtain that

$$S(H(Q), C_0) = S(G_{sf}(Q), K).$$

Adding C_1QC_1' to both sides of this last equation, and applying the function f_r, from Condition 2 we conclude that C_0 satisfies (26).

Now, suppose that there exists a controller $\mathscr{C}$ that satisfies condition (26). Define the following full information plant

$$\mathscr{G}_{fi}(Q) := \begin{cases} \dot{x} = Ax + QC_2'r + B_2u \\ v = C_1x + D_1u \\ y = [x'r']', \end{cases}$$

and let $G_{fi}(Q)$ denote its transfer matrix. Then, from the definition of $\mathscr{H}(Q)$ (cf. (24)), it follows that dynamic full information controller $C_{fi}(s) :=$

$[C(s)C_2\ I]$ satisfies

$$\mathscr{C}_{fi} \in \mathscr{A}(\mathscr{G}_{fi}(Q)) \quad \text{and} \quad f_r(C_1QC_1' + S(G_{fi}(Q), C_{fi})) < \gamma.$$

From Theorem 3.1 and Remark 3.1 it now follows that there exists

$$\mathscr{K} \in \mathscr{A}_m(\mathscr{G}_{sf}(Q)) \quad \text{such that} \quad f_r(C_1QC_1' + S(G_{sf}(Q), K)) < \gamma.$$

Hence, Condition 2 must hold and the proof is complete.

5. NUMERICAL SOLUTION OF THE GENERALIZED H_2 CONTROL PROBLEM

In this section we reformulate the convex program corresponding to the generalized H_2 control problem in a convenient form for finding numerical solutions. From the previous section it follows that it suffices to consider the state-feedback case. Below, we use the notation of Section 3, and to accommodate the case of output feedback, we will incorporate a positive semidefinite matrix R in the definition of the problem, see Remark 3.1 and Theorem 4.1.

Consider the plant $\mathscr{G}_{sf}$ introduced in (11). As shown in Theorem 3.2, and with the modification indicated in Remark 3.1, the generalized H_2 problem of finding a controller $\mathscr{K} \in \mathscr{A}_m(\mathscr{G}_{sf})$ such that $f_r(R + S(G_{sf}, K)) < \gamma$ may be reduced to the convex feasibility problem (15). Using well-established methods from the convex optimization literature, it is possible to develop effective numerical methods for solving this convex feasibility problem. See, for example, Boyd and Yang (1989) for the use of Kelley's cutting-plane method and Shor's subgradient method, in a convex problem similar to (15). However, in practice, it might be more advantageous to consider the following (perturbed) version of the generalized H_2 control problem.

Let $\beta > 0$ and $R = R' \geq 0$ be given, and consider the state-feedback plant

$$\mathscr{G}_{sf}(\beta) := \begin{cases} \dot{x} = Ax + [B_1\ \sqrt{\beta}I]w + B_2u \\ z = Cx + Du \\ y = x. \end{cases} \tag{27}$$

Find (if possible) a controller $\mathscr{K} \in \mathscr{A}_m(\mathscr{G}_{sf}(\beta))$ such that $f_r(R + S(G_{sf}(\beta), K)) < \gamma$.

The main advantage of adding a small tolerance β is that the (perturbed) generalized H_2 control problem may be solved using the convex program (16) instead of (15). Note that the convex program (16) has dimension qn, where q denotes the number of control inputs and n the state dimension. On the other hand, (15) involves a search over a space of dimension $qn + n(n-1)/2$. This significant reduction in the dimension of the search space is a consequence of the second part of Theorem 3.2. Note that, since the perturbed exogenous input matrix $[B_1\ \sqrt{\beta}I]$ has full row rank, the technical assumption in this part of the aforementioned theorem holds. Of course, if the matrices A, B_1, and B_2 satisfy this assumption, one can set $\beta = 0$.

It is obvious that, given any $\beta \geq 0$, $\mathscr{A}_m(\mathscr{G}_{sf}(\beta)) = \mathscr{A}_m(\mathscr{G}_{sf})$. Moreover, since we consider strict sub-optimality, the existence of a controller $\mathscr{K}$ that solves the problem with $\beta = 0$ guarantees the existence of a (sufficiently small) $\beta_{max} > 0$ such that, given any $\beta \in [0, \beta_{max}]$, $\mathscr{K}$ also solves the generalized H_2 problem for the perturbed plant $\mathscr{G}_{sf}(\beta)$. Thus, if we set β to a sufficiently small number (i.e. machine precision) we are certain that the two problems are equivalent for all practical purposes.

We now reduce the convex program (16), corresponding to the generalized H_2 problem for the plant (27), to a simpler program in which there are no equality constraints. Let $\beta > 0$ be given and consider the affine mapping $L(W, Y)$ defined in (13). Let (W_p, Y_p) denote a particular solution to the linear equation

$$L(W, Y) + \beta I = 0. \tag{28}$$

Note that such a solution always exists, for (A, B_2) is stabilizable. Moreover, (W_p, Y_p) may be found by computing a particular stabilizing gain $\mathscr{K}_p \in \mathscr{A}_m(\mathscr{G}_{sf}(\beta)) = \mathscr{A}_m(\mathscr{G}_{sf})$, and setting

$$Y_p = \text{controllability gramian of the pair } (A + B_2K_p, [B_1\ \sqrt{\beta}I]) > 0,$$

$$W_p := K_pY_p.$$

Let $\{(U_1, V_1), \ldots, (U_s, V_s)\}$, where $U_i = U_i' \in R^{n\times n}$ and $V_i \in R^{q\times n}$, denote a basis for the s-dimensional (generically $s = qn$) linear subspace

$$\{(W, Y) \in R^{q\times n} \times R^{n\times n} : Y = Y' \text{ and } AY + YA' + B_2W + W'B_2' = 0\}.$$

Let $a = [a_1\ a_2 \cdots a_s]' \in R^s$ be given. Define the affine mappings

$$\Psi(a) := Y_p + \sum_{i=1}^{s} a_iU_i \text{ and } \Phi(a) := W_p + \sum_{i=1}^{s} a_iV_i. \tag{29}$$

It now follows that (W, Y) satisfies $Y = Y'$ and $L(W, Y) + \beta I = 0$ if and only if there exists (a unique) vector $a \in R^s$ such that $Y = \Psi(a)$ and $W = \Phi(a)$.

Using Theorem 3.2, condition (16), we may conclude that there exists $\mathscr{K} \in \mathscr{A}_m(\mathscr{G}_{sf}(\beta))$ such that $f_r(R + S(G_{sf}(\beta), K)) < \gamma$ if and only if there

exists $a \in R^s$ such that

$$\begin{cases} \psi(a) := \lambda_{\max}(-\Psi(a)) < 0 \\ \phi_r(a) := f_r(R + M(\Phi(a), \Psi(a))) < \gamma, \end{cases} \tag{30}$$

where the mapping $M(.,.)$ was defined in (14). Furthermore, if a_{sol} solves (30), then $K_{sol} := \Phi(a_{sol})\Psi^{-1}(a_{sol})$ is a solution to our problem. It is important to note that the functions ψ and ϕ_r are convex on their domains of definition. Thus, (30) is a convex feasibility problem over R^s. This is a consequence of Theorem 3.3.

In some cases, one may be interested in computing the optimal generalized H_2 performance

$$v_r := \inf \{f_r(R + S(G_{sf}(\beta), K)) : \mathcal{K} \in \mathcal{A}_m(\mathcal{G}_{sf}(\beta))\},$$

and a (nearly) optimal controller. This can be accomplished by solving the convex optimization problem

$$v_r = \min \{\phi_r(a) : a \in R^s \text{ such that } \psi(a) < 0\}, \tag{31}$$

where $\psi(.)$ and $\phi_r(.)$ are defined in (30). Note that the constraint set $\Theta = \{a \in R^s : \psi(a) < 0\}$ is nonempty, for $0 \in \Theta$. Also, since $\beta > 0$, the infimum is attained in Θ. This is because, at a solution, the controllability gramian is positive.

Most methods for solving the convex problems (30) or (31) require not only the evaluation of the functions $\psi(.)$ and $\phi_r(.)$, but also the evaluation of subgradients for these functions. Closed form expressions for a subgradient of $\psi(a)$, at a point $a \in R^s$, and a subgradient of $\phi_r(a)$, at a point $a \in \Theta$, are easy to obtain using standard tools for calculating subgradients. A list of some of these tools may be found in Boyd and Barratt (1990). For brevity, details are omitted.

The algorithm we have implemented (algorithm GH2) to solve the convex programs (30) or (31) is a "deep cut" version of the "ellipsoid algorithm with constraints" described by Boyd and Barratt (1990). Roughly, this algorithm produces a sequence of volume decreasing ellipsoids in R^s that contain feasible points or optimizers. The algorithm GH2 determines in a finite number of iterations (provided that $\gamma \neq v_r$) whether the convex feasibility problem (30) has a solution. When solving the constrained optimization problem (31), GH2 determines in a finite number of steps an exit point that satisfies $\psi(a_{exit}) < 0$ and $\phi_r(a_{exit}) \leq v_r + \epsilon$, where ϵ is an arbitrary positive constant. Hence, the generalized H_2 cost with the controller $K_{exit} := \Phi(a_{exit})\Psi^{-1}(a_{exit})$ is within ϵ of the optimal generalized H_2 performance v_r. To run the algorithm GH2 an initial point $a_0 \in R^s$ and a symmetric positive definite matrix $A_0 \in R^{s \times s}$ must be specified. The pair (a_0, A_0) defines the initial ellipsoid $E_0 := \{a \in R^s : (a - a_0)' A_0^{-1} (a - a_0) \leq 1\}$. The initial ellipsoid E_0 should contain the feasible set defined by (30) to guarantee convergence (Boyd and Barratt, 1990). Even though it seems possible (based on the structure of our problem) to find (a_0, A_0) so that the initial ellipsoid E_0 contains the feasible set (30), we use $a_0 = 0$ and $A_0 = \kappa I$, where κ denotes some positive constant.

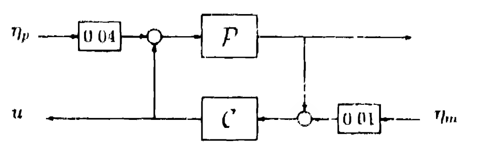


FIG. 3. The feedback configuration of the numerical example.

We conclude this section with a numerical example. This example is taken from Boyd and Barratt (1990). Consider the single-input–single-output feedback system shown in Fig. 3, where the plant is given by

$$P(s) = \frac{1}{s^2} \frac{10 - s}{10 + s}.$$

The controlled output vector is chosen to be $z = [e/\gamma_1 \; \dot{e}/\gamma_2 \; u/\gamma_3]'$, where e denotes the regulation error, u the control input, and the constant weights γ_1, γ_2, γ_3 will be specified later. The exogenous input vector is given by the process noise η_p, and the measurement noise η_m. That is, $w = [\eta_p \; \eta_m]'$.

Let T_{zw} denote the 3×2 closed loop transfer matrix from w to z corresponding to the controller $\mathscr{C}$, and define

$$S(C) = \frac{1}{2\pi} \int_{-\infty}^{\infty} T_{zw}(j\omega) T_{zw}^*(j\omega) \, d\omega.$$

First, we consider the problem of finding (if possible) an admissible controller $\mathscr{C}$ such that

$$f_r(S(C)) < 1, \quad r = 2 \text{ or } \infty. \tag{32}$$

The second synthesis problem is to compute the optimal generalized H_2 performance

$$v_r := \inf \{f_r(S(C)) : \mathscr{C} \text{ admissible}\}, \tag{33}$$

and a controller that comes as closed as desired to the optimal performance v_r. Two sets of constant weights are used for this example.

Specification 1. $\gamma_1 = 0.07$, $\gamma_2 = 0.04$, $\gamma_3 = 0.15$.

Specification 2. $\gamma_1 = 0.0698$, $\gamma_2 = 0.04$, $\gamma_3 = 0.149$.

Note that the two specifications are very close to each other. This is important to determine the sensitivity of our method.

Before performing any calculations there are a few important things to observe. First of all, due

TABLE 1. GENERALIZED H_2 COSTS OF LQG CONTROLLERS AND GENERALIZED H_2 COSTS OF FILTERING

Quantity	Specification 1	Specification 2
$\gamma_1, \gamma_2, \gamma_3$	0.07, 0.04, 0.15	0.0698, 0.04, 0.149
Trace $(S(C_{lqg}))$	2.1255	2.1333
$\lambda_{max}(S(C_{lqg}))$	1.3319	1.3341
$d_{max}(S(C_{lqg}))$	1.2297	1.2307
$\lambda_{max}(C_1QC_1')$	0.7371	0.7373
$d_{max}(C_1QC_1')$	0.7071	0.7071

the inequality

$$d_{max}(.) \le \lambda_{max}(.) \le \text{trace}\,(.),$$

it is obvious that if the standard LQG controller, say $\mathscr{C}_{lqg}$, satisfies trace $(S(C_{lqg})) < 1$, then this controller is also a solution to the feasibility problem (32). Secondly, the condition

$$f_r(C_1QC_1') < 1,$$

where C_1 is the output matrix corresponding to controlled output z, is necessary for the existence of a controller $\mathscr{C}$ that satisfies (32), see Theorem 4.1. The quantity $f_r(C_1QC_1')$, which is computed from the ARE (21), is coordinate independent and it represents the generalized H_2 "cost of filtering". The numerical values of the quantities just discussed are given in Table 1.

From Table 1 it follows that neither the LQG controller satisfies the specifications nor the costs of filtering violate the specifications. Thus, we have no choice other than running the ellipsoid algorithm GH2 to determine whether the specifications can be met.

The results of running the ellipsoid algorithm GH2 to solve the feasibility problem (32) are shown in Table 2. The parameter β introduced earlier was set to machine precision, in this case $\beta \approx 10^{-16}$. The particular solution to (28) was chosen to be the LQR gain. The initial ellipsoid was chosen according to the rule described earlier with a constant $\kappa = 1.4 \times 10^{-5}$.

In Table 2, ψ and ϕ_r are as defined in (30). Note that the number of subgradient evaluations for the objective function ϕ_r is much larger than those corresponding to the constraint function ψ. Thus, in this example, the algorithm spends most of its time iterating on the objective function ϕ_r. The quantities ϕ_r^{lb} and ϕ_r^{ub} denote the lower and upper bounds on the objective function ϕ_r, used to stop the algorithm. See Boyd and Barratt (1990) for further details on stopping criteria. For the case $r = \infty$, i.e. $f_r(.) = d_{max}(.)$, we can compare the results of Table 2 with those of Boyd and Barratt. Their method took 59 iterations with 47 evaluations of their cost function to determine that Specification 1 is feasible, while 34 iterations with 22 function evluations were required to find that Specification 2 is not feasible.

Table 3 shows the results of running GH2 to compute the optimal performance v_r with an absolute error of 10^{-5}. The parameter β and the initial ellipsoid are as before. On exit, the algorithm also gives a nearly optimal controller $\mathscr{C}_{exit}$ that satisfies $f_r(S(C_{exit})) \le v_r + 10^{-5}$. Of course, on exit, the algorithm just produces a state-feedback gain. The filter gain is precomputed according to Theorem 4.1. The row labeled improvement over LQG controller corresponds to the quantity

$$100 \times \sqrt{\frac{f_r(S(C_{lqg}))}{f_r(S(C_{exit}))}} - 100.$$

To give an idea of the possible differences between generalized H_2 synthesis, and standard

TABLE 2. RESULT OF RUNNING THE ALGORITHM GH2 TO DETERMINE THE FEASIBILITY OF THE SPECIFICATIONS

Quantity	Specification 1		Specification 2	
Norm (f_r)	λ_{max}	d_{max}	λ_{max}	d_{max}
Iterations	14	42	14	74
Subgrad. evaluations ψ	5	4	5	4
Subgrad. evaluations ϕ_r	9	38	9	70
Exit condition. $\psi < 0$ and	$\phi_2^{lb} > 1$	$\phi_\infty^{ub} < 1$	$\phi_2^{lb} > 1$	$\phi_\infty^{lb} > 1$
Feasible?	No	Yes	No	No

TABLE 3. RESULT OF RUNNING THE ALGORITHM GH2 TO COMPUTE THE OPTIMAL GENERALIZED H_2 COSTS

Quantity	Specification 1		Specification 2	
Norm (f_r)	λ_{max}	d_{max}	λ_{max}	d_{max}
Iterations	84	102	87	110
Subgrad. evaluations: ψ	5	4	5	4
Subgrad. evaluations: ϕ_r	79	98	82	106
$\sqrt{v_r}$	1.0835	0.9997	1.0848	1.0001
Improvement over LQG	6.51%	10.92%	6.47%	10.92%

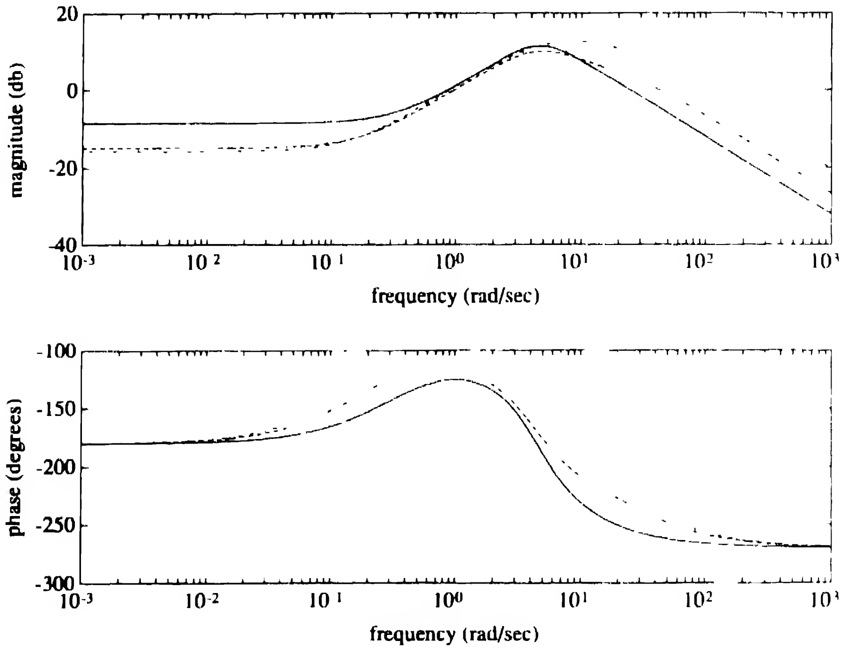


Fig. 4. Frequency responses of optimal controllers corresponding to Specification 1. Solid: LQG, dashed: $r = 2$, dash–dot: $r = \infty$

LQG synthesis, the nearly optimal generalized H_2 controllers generated by the ellipsoid algorithm, and the standard LQG controller, corresponding to Specification 1, are given below;

$$C_{lqg} = -24.9091 \times \frac{(s + 0.3441)(s + 10)}{(s + 3.1443 \pm j3.3804)(s + 10.8233)},$$

$$C_{ext,2} = -46.9815 \times \frac{(s + 0.1773)(s + 10)}{(s + 3.4866 \pm j2.7734)(s + 23.2865)}, \quad (\lambda_{max}),$$

$$C_{ext,\infty} = -96.8849 \times \frac{(s + 0.1447)(s + 10)}{(s + 6.1952 \pm j1.3494)(s + 21.4151)}, \quad (d_{max}).$$

The fact that the three controllers share a zero at $s = -10$ is not mere coincidence. This zero is fixed by the Kalman filter. Indeed, since the best estimate for the state of the "all-pass" factor is the model itself (i.e. the filtering gain for this state is zero) $s = -10$ is always a pole of the Kalman filter. Moreover, due to the special structure of this example, this pole becomes a zero of the controller for any state-feedback gain. The frequency responses of these controllers are shown in Fig. 4. Note that the three controllers perform more or less the same classical task—lead compensation. The main differences are the low-frequency gain, and the amount of phase lead introduced.

6 CONCLUSIONS

In this paper we have considered a (suboptimal) generalized H_2 control problem. This synthesis problem is well motivated since it represents a problem of disturbance attenuation, as measured by the peak-value of the controlled outputs, in the face of unknown energy exogenous inputs.

We have shown that when the state of $\mathcal{G}$ (or the state and the exogenous input) is available for feedback, static state-feedback gains offer the best possible performance for the generalized H_2 synthesis problem. That is, in the state-feedback case, dynamic compensators buy nothing extra as far as the generalized H_2 synthesis problem concerns. Further, necessary and sufficient conditions for the solvability of the generalized H_2 synthesis problem with state-feedback, as well as a formula for a solution, are obtained via a finite-dimensional convex program. In the case of output feedback, it is shown that generalized H_2 controllers can be chosen to be a combination of a standard Kalman filter, and a state-feedback gain that solves the generalized H_2 synthesis problem for a suitably constructed auxiliary plant. Due to the convexity properties of our approach, a solution to the generalized H_2 problem may be readily found.

The results in this paper may be extended in several ways. Firstly, they can be readily modified to solve a multicriterion generalized H_2 control problem. That is, the problem of finding a stabilizing controller such that several closed

loop transfer matrices are jointly small as measured by their generalized H_2 costs. These extensions are trivial. Secondly, the results in this paper may be combined with those in Khargonekar and Rotea (1991b) in order to solve the synthesis problem of finding a controller such that an "upper bound" for the generalized H_2 cost of one closed loop transfer matrix and the H_∞ norm of some other closed loop transfer matrix are below desirable levels. This multicriterion control problem is extremely important since it represents a problem of nominal performance, as measured by a generalized H_2 cost, with robust stability, as enforced by the H_∞ constraint. The solution to this problem may be found in Rotea and Khargonekar (1991). The extension of this mixed problem to accommodate more than one generalized H_2 performance measure is relatively straightforward. Extensions to discrete-time problems are also possible by combining the results in this paper with those in Kaminer *et al.* (1993).

Acknowledgements—This work was supported in part by the National Science Foundation under RIA no ECS-91-08493 The author is grateful to P. P. Khargonekar for reading the first draft of this manuscript and making suggestions for improving the clarity of the paper. The author is also grateful to S. P. Boyd for many interesting discussions regarding convex optimization.

REFERENCES

Bernussou, J., P. L. D. Peres and J. C. Geromel (1989). An LP oriented procedure for quadratic stabilization of uncertain systems. *Syst. Contr. Lett.*, **13**, 65–72.

Boyd, S. P. and C. H. Barratt (1990). *Linear Controller Design: Limits of Performance*. Prentice-Hall, NJ.

Boyd, S. P. and O. Yang (1989). Structured and simultaneous Lyapunov functions for system stability problems. *Int. J. Control*, **49**, 2215–2240.

Corless, M., G. Zhu and R. E. Skelton (1989). Robustness properties of covariance controllers. *Proc. of the 28th Conf. on Decision and Control*, Tampa, FL.

Doyle, J. C. (1984). ONR/Honeywell Workshop on Advances in Multivariable Control. Minneapolis, MN.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State-space solutions to standard H_2 and H_∞ and control problems. *IEEE Trans. Aut. Control*, **34**, 831–847.

Grimble, M. J. (1990). Relationship between the trace and the maximum eigenvalue norms for linear quadratic control design. *IEEE Trans Aut. Control*, **35**, 1176–1181.

Kalman, R. E. (1960). Contributions to the theory of optimal control. *Bul. Soc. Mat. Mexico*, **5**, 102.

Kaminer, I., P. P. Khargonekar and M. A. Rotea (1993). Mixed H_2/H_∞ control for discrete-time systems via convex optimization. *Automatica*, **29**, 57–70..

Khargonekar, P. P. and M. A. Rotea (1991a). Multiple objective control of linear systems: the quadratic norm case. *IEEE Trans. Aut. Control*, **36**, 14–24.

Khargonekar, P. P. and M. A. Rotea (1991b). Mixed H_2/H_∞ control: A convex optimization approach. *IEEE Trans. Aut. Control*, **36**, 824–837.

Khargonekar, P. P., I. R. Petersen and M. A. Rotea (1988). H_∞-optimal control with state-feedback. *IEEE Trans. Aut. Control*, **33**, 786–788.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

Packard, A. and G. Becker (1990). State-feedback solutions to quadratic stabilization. *Proc. of the 28th Annual Allerton Conf. on Communication, Control, and Computing*, pp. 768–769.

Rotea, M. A. (1991). The generalized H_2 control problem Technical report. School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN *Proc. of 1st IFAC Symposium on Design Methods of Control Systems*, Zurich, Vol. 1, pp. 112–117 (abridged).

Rotea, M. A. and P. P. Khargonekar (1991). Generalized H_2/H_∞ control via convex optimization. *Proc. of the 30th Conference on Decision and Control*, Brighton, U.K. Vol. 3, pp. 2719–2720.

Wilson, D. A. (1989). Convolution and Hankel operator norms for linear systems. *IEEE Trans. Aut. Control*, **34**, 94–97.

Wilson, D. A. (1990). Extended optimality properties of the linear quadratic regulator and stationary Kalman filter. *IEEE Trans Aut. Control*, **35**, 583–585.

Zhu, G. and R. E. Skelton (1991). Mixed L_2 and L_∞ problems by weight selection in quadratic optimal control. *Int J. Control*, **53**, 1161–1176.

Automatica, Vol 29, No 2, pp 387–400 1993
Printed in Great Britain

New Results on Composite Control of Singularly Perturbed Uncertain Linear Systems*

M. CORLESS,†‡ F. GAROFALO§ and L. GLIELMO§||

New results on the invariance of certain properties of singularly perturbed systems under output feedback are used in the construction of "stabilizing" controllers for a class of uncertain singularly perturbed systems whose fast dynamics are unstable.

Key Words—Singular perturbations, robust control, Lyapunov methods, nonlinear control systems

Abstract—We consider the problem of "stabilizing" uncertain, singularly perturbed, linear time-varying, systems whose fast dynamics can be unstable. We propose a class of nonlinear composite controllers which assure global uniform ultimate boundedness of the trajectories of the closed loop system, provided the singular perturbation parameter is sufficiently small. These controllers consist of a fast controller which stabilizes the fast dynamics and a slow controller which yields the desired stability properties for the slow dynamics. We consider the structure of the fast controller to be simpler than structures previously proposed in the literature. To obtain these controllers, we first develop some new results for singularly perturbed systems under output feedback. In particular, it is shown that the *matching assumption*, which deals with the manner in which the uncertainties enter the system, and is made on the reduced-order system, is invariant under linear feedback of the fast variable.

1 INTRODUCTION

UNCERTAIN MODEL parameters and multiple-time scale behavior of the plant under consideration are common problems encountered in the synthesis of control systems. Since it is not always possible to measure all the quantities which affect a plant, because of technological, economical or practical constraints, it is usual engineering practice to consider a nominal plant during the design of the control system, and then check *a posteriori* the satisfaction of the design specifications, either directly on the plant or via detailed numerical simulations. A similar procedure is often used when dealing with plants characterized by multiple-time scale behavior; in that case, only the slow dynamics are considered in the design stage and the full dynamics are taken into account in the validation phase.

In recent years, sounder mathematical foundations for these engineering practices have been sought, so that, while maintaining controller design sufficiently simple, some *a priori* knowledge of a more complex system than the nominal one can be embedded in the early stages of design. Good examples of this approach are the deterministic control of uncertain systems via Lyapunov's second method and singular perturbation theory.

In the former (Corless and Leitmann, 1988; Leitmann, 1989a contain a comprehensive list of references) the uncertain quantities are assumed to range in *a priori* known sets and to enter the dynamics of the plant via given subspaces, but, apart from some mild technical conditions, nothing is assumed to be known on their actual time history. Lyapunov's second method (Kalman and Bertram, 1960) is then used to construct stabilizing controllers.

Singular perturbation theory (Kokotović *et al.*, 1986; Kokotović and Khalil, 1986) permits the decomposition of a two-time scale system into two subsystems of lower order, one for each time scale. This, in turn, divides the full control system design into two parts and, moreover, yields an *a priori* evaluation of the robustness of the procedure, i.e. of how different from each

* Received 19 October 1990, revised 5 June 1992, received in final form 28 August 1992. A preliminary version of this paper was presented at the IFAC Symposium on Nonlinear Control Systems which was held in Capri, Italy during June 1989. The Published Proceedings of this IFAC Meeting may be ordered from Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U K. This paper was recommended for publication in revised form by Editor T Başar

† School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907, U S A

‡ Supported by the U S National Science Foundation under Grants MSM-87-06927 and MSS-90-57079

§ Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli "Federico II", via Claudio 21, I-80125 Napoli, Italy

|| This work was performed during a stay at the School of Aeronautics and Astronautics, Purdue University, supported by Consiglio Nazionale delle Ricerche, Italy, under Grant 203.07.17

other the time scales must be in order for this approximate design to work.

The combination of the two approaches in order to obtain robust controllers for a singularly perturbed system subject to uncertain parameter variations, has been studied in a number of papers. More specifically, in Leitmann *et al.*, (1986), Ryan and Corless (1991), Corless (1987, 1989), Leitmann (1989b) and Corless *et al.* (1990) the method of singular perturbations is used to assess the robustness of feedback control laws with respect to fast dynamics neglected in the design stage. In Garofalo (1988), Garofalo and Leitmann (1988), Corless *et al.* (1989), Garofalo and Glielmo (1989) and Garofalo and Leitmann (1990) the fast uncertain dynamics are not assumed to be stable, and the proposed controllers are composite controllers in the sense that they consist of a "fast" controller "composed" with a slow one; roughly speaking the fast controller stabilizes the fast dynamics and the slow controller yields the desired stability properties for the slow system dynamics. Similar ideas have been successfully used in the framework of adaptive control; see, e.g. Taylor *et al.* (1989).

In this paper, as in Corless *et al.* (1989), we consider a class of linear time-varying uncertain singularly perturbed systems, whose boundary-layer subsystem may be unstable, and we propose a class of nonlinear composite controllers which assure ultimate boundedness of the trajectories of the closed loop system for sufficiently small values of the singular perturbation parameter. We emphasize that, although the design philosophy is that of composite control, the structure of the fast controller is simpler than the structures previously proposed in the literature. This simpler structure is achieved by exploiting the linearity of the fast dynamics.

In order to develop our new controllers, we first obtain some new results on singularly perturbed systems subject to output feedback. In particular we show that, for the reduced-order system describing the slow dynamics, properties such as controllability, observability, and matching of uncertainties are invariant under well-posed output feedback laws.

The results are illustrated by application to a simple flexible mechanical system.

2 PROBLEM STATEMENT

We consider uncertain singularly perturbed systems described by

$$\dot{x}(t) = A_{11}(q(t))x(t) + A_{12}(q(t))z(t) + B_1(q(t))u(t) + b_1(q(t)), \tag{1a}$$

$$\mu\dot{z}(t) = A_{21}(q(t))x(t) + A_{22}(q(t))z(t) + B_2(q(t))u(t) + b_2(q(t)), \tag{1b}$$

where $t \in \mathcal{T} \triangleq [t_0, \infty)$ is the "time", $x(t) \in R^n$ and $z(t) \in R^m$ are state vectors, $u(t) \in R^p$ is the control input, $q(t) \in R^{\eta}$ is the vector of uncertainties, $A_{ij}(\cdot)$, $B_i(\cdot)$ and $b_i(\cdot)$ are continuous matrix functions of appropriate dimensions, and $\mu > 0$ is the singular perturbation parameter. Note that $b_1(q(\cdot))$ and $b_2(q(\cdot))$ can be interpreted as disturbance inputs. Regarding the uncertainty $q(\cdot)$ we make the following assumption.

Assumption 1. There exist a compact set $Q \subset R^{\eta}$ and a positive constant k_q, both known, such that $q: \mathcal{T} \to Q$, and q is continuously differentiable with $\|\dot{q}(t)\| \le k_q$ on $\mathcal{T}$.†

The next assumption deals with the dependence of some of the matrices of equation (1) on q.

Assumption 2. The functions $A_{21}(\cdot)$, $A_{22}(\cdot)$, $B_2(\cdot)$ and $b_2(\cdot)$ are continuously differentiable.

Note that we do not require continuous differentiability of all the $A_{ij}(\cdot)$ since we do not require an exact decomposition of the system dynamics into a slow and a fast part; compare this with Kokotovič *et al.* (1986).

We also assume that the singularly perturbed system (1) is in standard form (Kokotovič *et al.*, 1986) as follows.

Assumption 3. The matrix $A_{22}(q)$ is invertible on Q.

Letting $\mu = 0$ in equation (1), the associated reduced-order system is given by‡

$$\dot{x}_0 = A_0(q)x_0 + B_0(q)u + b_0(w), \tag{2a}$$

$$z_0 = h(x_0, u, q), \tag{2b}$$

with

$$h(x_0, u, q) \triangleq -A_{22}^{-1}(q)[A_{21}(q)x_0 + B_2(q)u + b_2(q)]. \tag{3}$$

In the above equations, $x_0 \in R^n$ is the slow time scale approximation of state x, $z_0 \in R^m$ is the "quasi steady state" component of the state z,

† If $\mathcal{S}$ is a set, by "on $\mathcal{S}$" we mean "for all $s \in \mathcal{S}$".

‡ In what follows, arguments of functions will be omitted when no confusion is likely to occur.

and

$$A_0(q) \triangleq A_{11}(q) - A_{12}(q)A_{22}^{-1}(q)A_{21}(q), \quad (4a)$$

$$B_0(q) \triangleq B_1(q) - A_{12}(q)A_{22}^{-1}(q)B_2(q), \quad (4b)$$

$$b_0(q) \triangleq b_1(q) - A_{12}(q)A_{22}^{-1}(q)b_2(q). \quad (4c)$$

We introduce the following definition.

Definition 1. A system of the form (2) has uncertainties matched with the input around a pair $(\bar{A}_0, \bar{B}_0)$, $\bar{A}_0 \in R^{n\times n}$, $\bar{B}_0 \in R^{n\times p}$ if

- the pair $(\bar{A}_0, \bar{B}_0)$ is stabilizable;
- there exist matrix functions E, G, g, with $G(q)$ nonsingular on Q, s.t.

$$A_0(q) = \bar{A}_0 + \bar{B}_0 E(q),$$
$$B_0(q) = \bar{B}_0 G(q),$$
$$b_0(q) = \bar{B}_0 g(q),$$

on Q.

Regarding the reduced-order system we assume the following.

Assumption 4 (uncertainty matching assumption). There exists a pair $(\bar{A}_0, \bar{B}_0)$, $\bar{A}_0 \in R^{n\times n}$, $\bar{B}_0 \in R^{n\times p}$, s.t. system (2) has uncertainties matched with the input around $(\bar{A}_0, \bar{B}_0)$.

It is worth noting that in Garofalo and Leitmann (1990) and Garofalo and Glielmo (1989), matching of the uncertainties is assumed on each of the two subsystems (1a) and (1b); this is different from here.

In this paper we consider the problem of "stabilizing" system (1) via a memoryless state feedback controller. Since system (1) is uncertain, the controller must assure the stability property independently of the particular form the uncertainty $q(\cdot)$ takes. The *continuous* controllers we propose guarantee uniform ultimate boundedness of the trajectories of the closed loop system for all $\mu > 0$ sufficiently small. Loosely speaking this means that the trajectories will be uniformly bounded and, after a sufficiently long time, be confined to a ball containing the origin of the state space.

We point out that we do not assume the matrix $A_{22}(q)$ to be stable on Q; in other words, the boundary-layer system† associated with the full order system (1) is not assumed to be asymptotically stable. This, from a composite control point of view, implies that a controller must be designed to stabilize the boundary layer system. Usually this is accomplished by feeding back the boundary layer state $z - z_0$. This is not possible here since z_0 depends on the uncertainty q. However we note that, since the fast subsystem is linear, the boundary layer system can be made asymptotically stable by simply feeding back the state z. This type of controller, however, changes the reduced-order system.

In Section 4 we analyse the effects of linear feedback of the variable z on the reduced-order system and on the matching assumption. To do so, we first develop some new results on singularly perturbed systems under linear output feedback; these are presented in the next section.

† A precise definition is given in the next section.

3 LINEAR OUTPUT FEEDBACK AND THE REDUCED-ORDER SYSTEM

Output feedback control of singular perturbed systems has been considered by many authors. Some of them (Chemouil and Wahdan, 1980; Fossard and Magni, 1980; Khalil, 1985; Moerder and Calise, 1985, 1988; Calise *et al.*, 1990) have studied the design of output feedback controllers for singularly perturbed systems, others (Khalil, 1981, 1984; Vidyasagar, 1985a, b; Corless and Glielmo, 1993) have dealt with the robustness of output feedback controllers designed on the basis of the reduced-order system (i.e. for $\mu = 0$) but applied to the full order system (for $\mu > 0$).

Here, we analyse the effects on the reduced-order system, of a linear output feedback applied to the full order system.

Suppose system (1) is endowed with a measured output of the form

$$y = C_1(q)x + C_2(q)z + c(q), \quad (5)$$

where $y(t) \in R^l$, and $C_1(\cdot)$, $C_2(\cdot)$ and $c(\cdot)$ are continuous matrix functions.

If system (1), (5) is subject to a linear output feedback controller of the form

$$u(t) = Fy(t) + \bar{u}(t), \quad (6)$$

then the resulting closed loop system is

$$\dot{x} = \bar{A}_{11}(q)x + \bar{A}_{12}(q)z + B_1(q)\bar{u} + \bar{b}_1(q), \quad (7a)$$

$$\mu\dot{z} = \bar{A}_{21}(q)x + \bar{A}_{22}(q)z + B_2(q)\bar{u} + \bar{b}_2(q), \quad (7b)$$

with

$$\bar{A}_{ij} \triangleq A_{ij} + B_i F C_j, \quad (8a)$$

$$\bar{b}_i \triangleq b_i + B_i F c, \quad (8b)$$

for $i, j = 1, 2$.

Clearly, system (7) is in standard form, i.e. the equation

$$0 = \bar{A}_{21}(q)x_0 + \bar{A}_{22}(q)z_0 + B_2(q)\bar{u} + \bar{b}_2(q),$$

has a unique solution z_0, if and only if the matrix $\bar{A}_{22}(q)$ is nonsingular on Q. Under this

hypothesis we have

$$z_0 = -\tilde{A}_{22}^{-1}(q)[\tilde{A}_{21}(q)x_0 + B_2(q)\bar{u} + \tilde{b}_2(q)]. \quad (9)$$

Substituting (9) into (7) we obtain the reduced-order system

$$\dot{x}_0 = \tilde{A}_0(q)x_0 + \tilde{B}_0(q)\bar{u} + \tilde{b}_0(q), \quad (10a)$$

$$y = \tilde{C}_0(q)x_0 + \tilde{D}_0(q)\bar{u} + \tilde{c}_0(q), \quad (10b)$$

with

$$\tilde{A}_0 \triangleq \tilde{A}_{11} - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{A}_{21}, \quad (11a)$$

$$\tilde{B}_0 \triangleq B_1 - \tilde{A}_{12}\tilde{A}_{22}^{-1}B_2, \quad (11b)$$

$$\tilde{b}_0 \triangleq b_1 - \tilde{A}_{12}\tilde{A}_{22}^{-1}\tilde{b}_2, \quad (11c)$$

$$\tilde{C}_0 \triangleq C_1 - C_2\tilde{A}_{22}^{-1}\tilde{A}_{21}, \quad (11d)$$

$$\tilde{D}_0 \triangleq -C_2\tilde{A}_{22}^{-1}B_2, \quad (11e)$$

$$\tilde{c}_0 \triangleq c - C_2\tilde{A}_{22}^{-1}\tilde{b}_2. \quad (11f)$$

We introduce the following definition.

Definition 2. The reduced-order system (10) is well-defined if the matrix $\tilde{A}_{22}(q)$ is nonsingular on Q.

System (10) is obtained by first applying the output feedback law (6) to system (1), (5) and then letting $\mu = 0$. We now reverse the order of these operations by applying controller (6) to the original reduced-order system associated with (1), (5). This is given by

$$\dot{x}_0 = A_0(q)x_0 + B_0(q)u + b_0(q), \quad (12a)$$

$$y = C_0(q)x_0 + D_0(q)u + c_0(q), \quad (12b)$$

where A_0, B_0, and b_0 are given by (4), and

$$C_0 \triangleq C_1 - C_2A_{22}^{-1}A_{21}, \quad (13a)$$

$$D_0 \triangleq -C_2A_{22}^{-1}B_2, \quad (13b)$$

$$c_0 \triangleq c - C_2A_{22}^{-1}b_2. \quad (13c)$$

If system (12) is subject to controller (6), then

$$y = C_0(q)x_0 + D_0(q)Fy + D_0(q)\bar{u} + c_0(q). \quad (14)$$

For the above equation to have a unique solution for y, we require $I_l - D_0F$ to be invertible on Q. Under this hypothesis we can solve equation (14) for y and obtain the following closed loop system

$$\dot{x}_0 = \hat{A}_0(q)x_0 + \hat{B}_0(q)\bar{u} + \hat{b}_0(q), \quad (15a)$$

$$y = \hat{C}_0(q)x_0 + \hat{D}_0(q)\bar{u} + \hat{c}_0(q), \quad (15b)$$

where

$$\hat{C}_0 \triangleq (I_l - D_0F)^{-1}D_0, \quad (16a)$$

$$\hat{D}_0 \triangleq (I_l - D_0F)^{-1}D_0, \quad (16b)$$

$$\hat{A}_0 \triangleq A_0 + B_0F\hat{C}_0, \quad (16c)$$

$$\hat{B}_0 \triangleq B_0(I_p + F\hat{D}_0), \quad (16d)$$

$$\hat{c}_0 \triangleq (I_l - D_0F)^{-1}c_0, \quad (16e)$$

$$\hat{b}_0 \triangleq b_0 + B_0F\hat{c}_0. \quad (16f)$$

The following definition is then justified.

Definition 3. The closed loop system (15) is well-defined if $I_l - D_0(q)F$ is nonsingular on Q.

Hence, systems (10) and (15) are closely related: the first one is obtained by computing a reduced-order system after applying the output feedback control law (6); the second one is obtained by closing the loop directly on the reduced-order system (12). In other words, the operations "closing the loop" and "letting $\mu = 0$" are applied in different orders. The natural question arising is whether they commute: although the answer is intuitively affirmative, it is not obvious. We formalize it in the next theorem.

Theorem 1. Suppose Assumption 3 holds and consider systems (10) and (15).
(1) If one of them is well-defined, the other one is well-defined.
(2) They are identical.

Proof. System (10) is well-defined iff $\tilde{A}_{22}$ is nonsingular, and (15) is well-defined iff $I_l - D_0F = I_l + C_2A_{22}^{-1}B_2F$ is nonsingular. Actually, these conditions are equivalent, as can be readily proven using the matrix equality (see for example Kailath (1980))

$$\det(I + AB) = \det(I + BA), \quad (17)$$

where A and B are any matrices of suitable dimensions. Indeed, one has

$$\begin{aligned}\det(I_l - D_0F) &= \det(I_l + C_2A_{22}^{-1}B_2F)\\ &= \det(I_m + A_{22}^{-1}B_2FC_2)\\ &= \det(A_{22}^{-1})\det(\tilde{A}_{22}),\end{aligned} \quad (18)$$

hence $I_l - D_0F$ is invertible iff $\tilde{A}_{22}$ is invertible.

As regards the identity of the two systems, this is equivalent to the identity of corresponding "tilde" and "hat" matrices in (10) and (15). To demonstrate the latter, we first notice that

$$\begin{aligned}\tilde{A}_{12}\tilde{A}_{22}^{-1} - A_{12}A_{22}^{-1} &= (\tilde{A}_{12} - A_{12}A_{22}^{-1}\tilde{A}_{22})\tilde{A}_{22}^{-1}\\ &= B_0FC_2\tilde{A}_{22}^{-1},\end{aligned} \quad (19)$$

and, similarly,

$$\tilde{A}_{22}^{-1}\tilde{A}_{21} - A_{22}^{-1}A_{21} = \tilde{A}_{22}^{-1}B_2FC_0, \quad (20)$$

$$\tilde{A}_{22}^{-1}\tilde{b}_2 - A_{22}^{-1}b_2 = \tilde{A}_{22}^{-1}B_2Fc_0. \quad (21)$$

From this, equations (11) yield

$$\tilde{B}_0 = B_0(I_p - FC_2\tilde{A}_{22}^{-1}B_2), \quad (22a)$$

$$\tilde{C}_0 = (I_l - C_2\tilde{A}_{22}^{-1}B_2F)C_0, \quad (22b)$$

$$\tilde{A}_0 = A_0 + \bar{B}_0 F C_0 \tag{22c}$$
$$= A_0 + B_0 F \tilde{C}_0, \tag{22d}$$
$$\tilde{c}_0 = (I_l - C_2 \tilde{A}_{22}^{-1} B_2 F) c_0, \tag{22e}$$
$$\tilde{b}_0 = b_0 + B_0 F \tilde{c}_0 \tag{22f}$$
$$= b_0 + \bar{B}_0 F c_0. \tag{22g}$$

On the other hand, the matrix inversion formula (Kailath, 1980) gives

$$(I_l - D_0 F)^{-1} = I_l - C_2 \tilde{A}_{22}^{-1} B_2 F, \tag{23a}$$
$$I_p + F\hat{D}_0 = (I_p - FD_0)^{-1} \tag{23b}$$
$$= I_p - F C_2 \tilde{A}_{22}^{-1} B_2. \tag{23c}$$

By applying the above equalities to (16) and comparing with equations (22) and (11e), one can derive the desired result.

In view of Theorem 1, we shall refer only to system (10) and call it the reduced-order closed loop system. We will call system (2) the reduced-order open loop system. Moreover, we can introduce the following definition.

Definition 4. The output feedback control law (6) is well-posed for the singularly perturbed system (1), (5) if the reduced-order closed loop system is well-defined.

Remark 1. In view of Theorem 1, the well-posedness of a feedback law is equivalent to the requirement that one (hence all) of the matrices $\tilde{A}_{22}$, $I_l - D_0 F$, $I_p - FD_0$ be nonsingular on Q. Then, the useful formulas (23) apply.

Remark 2. Since corresponding "hat" and "tilde" matrices are equal, we can use either (16) or (22). The form and symmetry of these equations suggest some interpretations. Looking at system (10) from the control input point of view, we write

$$\bar{B}_0 = B_0 (I_p - FD_0)^{-1}, \tag{24a}$$
$$\tilde{A}_0 = A_0 + \bar{B}_0 F C_0, \tag{24b}$$
$$\tilde{b}_0 = b_0 + \bar{B}_0 F c_0. \tag{24c}$$

Thus, the effects of a well-posed output feedback law on the original reduced-order system (2) are equivalent to

- right-multiplying the input matrix by the nonsingular matrix $(I_p - FD_0)^{-1}$;
- feeding back, through the new input matrix, the quantity

$$F(C_0 x_0 + c_0) = F[C_1 x_0 + C_2 h(x_0, 0, q) + c].$$

A dual interpretation, from the output point of view, is obtained by considering

$$\tilde{C}_0 = (I_l - D_0 F)^{-1} C_0, \tag{25a}$$
$$\tilde{c}_0 = (I_l - D_0 F)^{-1} c_0, \tag{25b}$$
$$\tilde{A}_0 = A_0 + B_0 F \tilde{C}_0, \tag{25c}$$
$$\tilde{b}_0 = b_0 + B_0 F \tilde{c}_0. \tag{25d}$$

A further consequence of equations (24)–(25) is the following theorem.

Theorem 2. Suppose system (1), (5) is time-invariant (i.e. q is a constant function), satisfies Assumption 3, and the output feedback law (6) is well-posed. Then, the reduced-order closed loop system is stabilizable (detectable) iff the reduced-order open loop system is stabilizable (detectable).

Returning now to Assumption 4 of the previous section, we notice that, using (24), one obtains

$$\bar{B}_0 = \bar{B}_0 G (I_p - FD_0)^{-1} \triangleq \bar{B}_0 G_F, \tag{26a}$$
$$\tilde{A}_0 = \bar{A}_0 + \bar{B}_0 (E + G_F F C_0) \triangleq \bar{A}_0 + \bar{B}_0 E_F, \tag{26b}$$
$$\tilde{b}_0 = \bar{B}_0 (g + G_F F c_0) \triangleq \bar{B}_0 g_F, \tag{26c}$$

where G_F, E_F and g_F depend on q. We point out that, if the output feedback law is well-posed, the matrix $G_F(q)$ in (26a) is nonsingular on Q. The above equations lead to the following theorem which, for the particular scope of this paper, can be considered the main result of this section.

Theorem 3. Consider system (1), (5) and suppose Assumption 3 holds and the linear output feedback law (6) is well-posed. Then, the reduced-order closed loop system has uncertainties matched with the input around the pair $(\bar{A}_0, \bar{B}_0)$ iff the open loop reduced-order system has uncertainties matched with the input around the same pair.

4. STABILIZATION OF THE BOUNDARY-LAYER SYSTEM

At the end of Section 2 we mentioned that our intention is to use linear feedback of the fast variable z to stabilize the boundary-layer system. To this end, consider a feedback law of the form

$$u(t) = Fz(t) + \bar{u}(t), \tag{27}$$

where the control $\bar{u}$ is to be used in the design of a controller for the reduced-order system. Clearly, (27) is a special case of (5)–(6) with

$$l = m, \tag{28a}$$
$$c(q) \equiv 0, \tag{28b}$$

$$C_1(q) \equiv 0, \tag{28c}$$

$$C_2(q) \equiv I_m. \tag{28d}$$

Thus, we can utilize the results of the previous section.

In particular one has

$$G_F = G(I_p + FA_{22}^{-1}B_2)^{-1}, \tag{29}$$

$$\bar{B}_0 = B_0(I_p + FA_{22}^{-1}B_2)^{-1}$$
$$= \bar{B}_0 F_G, \tag{30}$$

$$C_0 = -A_{22}^{-1}A_{21}, \tag{31}$$

$$\bar{A}_{22} = A_{22} + B_2 F. \tag{32}$$

Consider now any nonsingular matrix $M \in R^{p \times p}$ and define a new input vector

$$v \triangleq M^{-1}\bar{u}; \tag{33}$$

the input matrix of the reduced-order closed loop system relative to this new input is

$$\bar{B}_0(q)M = \bar{B}_0 L(q), \tag{34}$$

with

$$L \triangleq G(I_p + F\bar{A}_{22}^{-1}B_2)^{-1}M$$
$$= G_F M. \tag{35}$$

We note that the matrix $L(q)$ depends on matrices F and M. These are specified in the next assumption.

Assumption 5. There exist a z-feedback matrix F and a matrix M s.t. on Q,

(1) $\bar{A}_{22}(q)$ is asymptotically stable,

(2) $L(q)$ is positive definite.

Roughly speaking, Assumption 5.1 assures us that (a) the feedback law (27) is well-posed, and (b) the boundary-layer system associated with the singularly perturbed system (7), (8) and (28) is asymptotically stable in the fast time scale $\tau = (t - t_0)/\mu$ for any uncertainty $q(\cdot)$ and any "frozen" time instant t_0. Of course, a stronger condition than mere stability could be chosen, such as constraining the eigenvalues of $\bar{A}_{22}(q)$ to a prescribed sector of the left half complex plane, thus ensuring sufficient damping of the fast time scale dynamics.

Assumption 5.2 will enable us to establish a feedback control law generating v (essentially a feedback "direction") which stabilizes the reduced-order system. In order for Assumption 5.2 to be satisfied a necessary condition is that the matrix $G_F(q)$ be nonsingular on Q; this is guaranteed by Theorem 3.

Remark 3. If Assumption 5 is assured with some matrices F and M_0, then a whole class of matrices M will assure Assumption 5.2 and some sort of optimization turns out to be possible. One possibility is that M be chosen so as to solve the following optimization problem:

$$\max_{M \in \mathcal{M}} \min_{q \in Q} \lambda_{\min}[L(q) + L(q)^T], \tag{36}$$

where $\mathcal{M}$ is defined as

$$\mathcal{M} \triangleq \{M \in R^{p \times p} : \|M\| = 1\}.$$

It will be clear in the following that this contributes to decreasing the gains of the proposed controllers.

Remark 4. If $G(q)$ is positive definite on Q, and $A_{22}(\cdot)$ and $B_2(\cdot)$ are constant over Q, i.e. they are not affected by uncertainties, then Assumption 5.2 is certainly satisfied for some M. In particular a "natural" choice for M is

$$M = I_p + F\bar{A}_{22}^{-1}B_2. \tag{37}$$

With this choice $L(q) \equiv G(q)$.

Remark 5. In the scalar input case, i.e. $p = 1$, and $G(q) > 0$ on Q, Assumption 5.2 requires essentially that the scalar $1 + F\bar{A}_{22}^{-1}(q)B_2(q)$ does not change sign over Q. In this case, $1 + F\bar{A}_{22}^{-1}B_2 = \det(I_p + F\bar{A}_{22}^{-1}B_2) = \det(A_{22}^{-1})\det(\bar{A}_{22})$ (employing (17)); hence, noting that $\det[\bar{A}_{22}(q)]$ does not change sign over q, since $\bar{A}_{22}(q)$ is asymptotically stable on Q in view of Assumption 5.1.† we obtain that a necessary and sufficient condition for Assumption 5.2 to hold is that $\det[A_{22}(q)]$ does not change sign over Q. This, in turn, is guaranteed by Assumption 3, provided the set Q is connected.

Remark 6. We notice that, in view of (28) equation (9) can be written as

$$z_0 = (I_m + A_{22}^{-1}(q)B_2(q)F)^{-1}h(x_0, \bar{u}, q), \tag{38}$$

from which we see that the effect of z-feedback on the slow part of the z state is equivalent to multiplication by the "gain" matrix $(I_m + A_{22}^{-1}B_2F)^{-1}$.

Since z-feedback modifies the reduced-order system (see equation (26)), it is desirable to reconsider the choice of the nominal dynamical matrix $\bar{A}_0$ of the reduced-order system, so as to minimize the magnitude of the uncertain terms.

From equation (26b) we have

$$\bar{A}_0(q) = \bar{A}_{K^*} + \bar{B}_0 E_{K^*}(q), \tag{39}$$

where

$$\bar{A}_{K^*} \triangleq \bar{A}_0 + \bar{B}_0 K^*, \tag{40a}$$

$$E_{K^*}(q) \triangleq E_F(q) - K^*, \tag{40b}$$

and the matrix $K^* \in R^{p \times n}$ solves the following

† $\det(-\bar{A}_{22}(q))$ is equal to the constant term of the characteristic polynomial of $\bar{A}_{22}(q)$.

problem

$$\min_{K \in R^{p \times n}} \max_{q \in Q} \|E_F(q) - K\|. \tag{41}$$

The stabilizability of the pair $(\bar{A}_0, \bar{B}_0)$ is equivalent to that of $(\bar{A}_{K^*}, \bar{B}_0)$.

5. A NONLINEAR CONTROLLER

In this section we utilize the properties of linear z-feedback to propose a nonlinear composite controller which guarantees ultimate boundedness for the trajectories of system (1).† For the sake of readability, we first summarize in the following the form of the matching assumption for the reduced-order system after z-feedback:

$$\bar{A}_0(q) = \bar{A}_{K^*} + \bar{B}_0 E_{K^*}(q), \tag{42a}$$

$$\bar{B}_0(q)M = \bar{B}_0 L(q), \tag{42b}$$

$$\bar{b}_0(q) = \bar{B}_0 g_f(q). \tag{42c}$$

The controller has the form suggested in (27),

$$u(t) = Fz(t) + \bar{u}(t), \tag{43}$$

and we now specify

$$\bar{u}(t) = Mw(x(t)), \tag{44}$$

thus realizing feedback of the slow state x. In equations (43) and (44) matrices F and M are those described in Assumption 5; the function w is specified in the following subsection.

5.1. *Structure of the nonlinear controller*

Choose any positive definite symmetric $Q_0 \in R^{n \times n}$, let P_0 be the unique positive definite symmetric solution of the Riccati equation

$$\bar{A}_{K^*}^T P_0 + P_0 \bar{A}_{K^*} + Q_0 - 2P_0 \bar{B}_0 \bar{B}_0^T P_0 = 0, \tag{45}$$

and define the functions

$$v(x) \triangleq \bar{B}_0^T P_0 x, \tag{46a}$$

$$s(v) \triangleq (1 + \|v\|)^{-1} v. \tag{46b}$$

Then, the function w has the form

$$w(x) = -\beta_1 v - \beta_2 s(\gamma v), \tag{47}$$

where, for the sake of simplicity, we have dropped the dependence of v on x. In equation (47), β_1 and β_2 are non-negative constants, γ is a positive constant. The control law is made up of two parts: (a) the linear part counteracts the effect of the uncertainties in the matrix $\bar{A}_0$; (b) the nonlinear part is a continuous approximation of a discontinuous law (for details see Gutman, 1979; Corless and Leitmann, 1981; Ambrosino *et al.*, 1985) and counteracts the disturbance input $\bar{b}_0$.

To choose β_1 and β_2, define

$$k_{F_{K^*}} \triangleq \max_{q \in Q} \|E_{K^*}(q)\|, \tag{48a}$$

$$k_l \triangleq (1/2) \min_{q \in Q} \lambda_{\min}[L(q) + L(q)^T], \tag{48b}$$

$$k_{g_f} \triangleq \max_{q \in q} \|g_F(q)\|, \tag{48c}$$

and

$$\beta_1 \triangleq k_l^{-1} + \frac{k_{F_{K^*}}^2}{2\lambda_{\min}(Q_0)k_l}, \tag{49a}$$

$$\beta_2 \triangleq k_{g_f}/k_l, \tag{49b}$$

and let

$$\beta_1 > \beta_1, \tag{50a}$$

$$\beta_2 \geq \beta_2. \tag{50b}$$

The only requirement on γ is that it be strictly positive. Nevertheless, increasing γ reduces the size of the ball of ultimate boundedness of the reduced-order system, and thus it may be desirable to select it as large as practically possible.

5.2. *A stability result*

For the control law given above it is possible to prove the following theorem.

Theorem 4. Consider system (1) subject to the nonlinear controller (43)–(50). Then there exists a constant $\mu^* > 0$ such that, whenever $\mu < \mu^*$, the trajectories of the closed loop system are uniformly ultimately bounded.

Remark 7. We would like to point out that the control design procedure presented in this paper is different from the usual composite control design procedure. In that approach (see Kokotović *et al.* (1986), for the general philosophy and, for example, Garofalo and Leitmann (1990) for application to uncertain systems) the slow control design is carried out on the basis of the original reduced-order system; the z-feedback is designed to be inactive when $z(t) = z_0(t)$, because it feeds back $z - z_0$.‡ Feeding back the state z, while being simpler, modifies z_0, as shown in equations (9) and (38) and, in turn, the reduced-order system. In a linear "certain" system, these effects can be

† See the Appendix for a definition of ultimate boundedness and for a lemma on how to guarantee this property.

‡ In the case of an uncertain singularly perturbed system, $z_0(\cdot)$ turns out to be dependent on the uncertainty $q(\cdot)$ (see equation (2b)). A nominal quasi steady state trajectory corresponding to a nominal uncertainty $\bar{q}(\cdot)$ is then considered.

canceled out by the x-feedback; in the present case, due to the presence of the uncertainties, these effects are partly embedded in the nominal reduced-order system.

Proof of Theorem 4. As in Saberi and Khalil (1984) the proof is based on a Lyapunov function for the closed loop system which is obtained as a convex combination of two Lyapunov functions, one for the slow and one for the fast time scale approximation of the system.

With the notation introduced in Sections 3 and 4 (assuming $C_1 \equiv 0$, $C_2 \equiv I_m$, $c \equiv 0$), the closed loop system can be written in the form

$$\dot{x} = \bar{A}_0(q)x - \bar{B}_0(q)M[\beta_1 v + \beta_2 s(\gamma v)] + \bar{b}_0(q) + \bar{A}_{12}(q)(z - H(x, q)), \quad (51a)$$

$$\mu \dot{z} = \bar{A}_{22}(q)(z - H(x, q)), \quad (51b)$$

where

$$H(x, q) \triangleq -\bar{A}_{22}^{-1}(q)[A_{21}(q)x + B_2(q)Mw(x) + b_2(q)]. \quad (52)$$

Consider now the positive definite quadratic function

$$V(x) \triangleq x^T P_0 x, \quad (53)$$

and evaluate its derivative along the solutions of system (51). Using equations (42) and (45) we have

$$\begin{aligned}\dot{V}(x, z, t) &= 2x^T P_0\{\bar{A}_{K^*}x - \bar{B}_0 L[\beta_1 v + \beta_2 s(\gamma v)] \\ &\quad + \bar{B}_0(E_{K^*}x + g_F)\} \\ &\quad + 2x^T P_0 \bar{A}_{12}(z - H) \\ &\leq -\lambda_{\min}(Q_0)\,\|x\|^2 - 2\beta_1 k_l\,\|v\|^2 \\ &\quad - 2\gamma^{-1}\beta_2 k_l \frac{\|\gamma v\|^2}{1 + \|\gamma v\|} + 2\,\|v\|^2 \\ &\quad + 2k_{E_{K^*}}\,\|v\|\,\|x\| + 2k_{g_F}\,\|v\| \\ &\quad + p_{xz}\,\|x\|\,\|z - H\|, \end{aligned} \quad (54)$$

where

$$p_{xz} \triangleq 2 \max_{q \in Q} \|P_0 \bar{A}_{12}(q)\|. \quad (55)$$

Using (50) we obtain

$$\dot{V}(x, z, t) \leq -p_{xx}\,\|x\|^2 + p_1 + p_{xz}\,\|x\|\,\|z - H\|, \quad (56)$$

with

$$p_{xx} \triangleq \lambda_{\min}(Q_0) - \frac{k_{E_{K^*}}^2}{2(\beta_1 k_L - 1)}, \quad (57a)$$

$$p_1 \triangleq 2\gamma^{-1}\beta_2 k_L. \quad (57b)$$

Consider now the q-dependent solution $S(q)$ of the Lyapunov equation

$$\bar{A}_{22}(q)^T S(q) + S(q)\bar{A}_{22}(q) = -I_m. \quad (58)$$

In view of the first part of Assumption 5 this solution is positive definite on Q. Moreover, since $\bar{A}_{22}(\cdot)$ is bounded as a result of the continuity of $\bar{A}_{22}(\cdot)$ and the compactness of Q, $S(\cdot)$ is also bounded. Differentiating equation (58) with respect to q_i, for $i = 1, 2, \ldots, \eta$, we obtain

$$\bar{A}_{22}(q)^T \frac{\partial S(q)}{\partial q_i} + \frac{\partial S(q)}{\partial q_i}\bar{A}_{22}(q) = -\left[\frac{\partial \bar{A}_{22}(q)^T}{\partial q_i} S(q) + S(q)\frac{\partial \bar{A}_{22}(q)}{\partial q_i}\right],$$
$$\text{for } i = 1, 2, \quad (59)$$

As a consequence of part 1 of Assumption 5, the solutions $\partial S(q)/\partial q_i$ of equation (59) exist, and they are bounded since $\bar{A}_{22}(\cdot)$ and $\partial \bar{A}_{22}(\cdot)/\partial q_i$ are bounded by Assumption 2, and $S(\cdot)$ is bounded. From this and from Assumption 1, it follows that finite positive constants k_{Sq_i}, $i = 1, 2, \ldots, \eta$, exist such that

$$\left\|\frac{\partial S(q(t))}{\partial q_i}\dot{q}_i(t)\right\| \leq k_{Sq_i}, \quad \begin{cases}\text{for any } q(\cdot) \\ \text{and any } t \in \mathcal{T}.\end{cases} \quad (60)$$

Let us now consider

$$W(x, z, t) \triangleq [z - H(x, q(t))]^T S(q(t)) \times [z - H(x, q(t))]; \quad (61)$$

evaluating its derivative along the solutions of closed loop system (51), we obtain

$$\begin{aligned}\dot{W}(x, z, t) &= \frac{\partial W}{\partial x}\dot{x} + \frac{\partial W}{\partial z}\dot{z} + \frac{\partial W}{\partial t} \\ &= -2(z - H)^T S J_x \dot{x} \\ &\quad + 2\mu^{-1}(z - H)^T S \bar{A}_{22}(z - H) \\ &\quad - 2(z - H)^T S J_q \dot{q} \\ &\quad + \sum_{i=1}^{\eta}(z - H)^T \frac{\partial S}{\partial q_i}(z - H)\dot{q}_i. \end{aligned} \quad (62)$$

In the previous equation J_x and J_q are the Jacobian matrices $J_x \triangleq \partial H/\partial x$ and $J_q \triangleq \partial H/\partial q$. The first one can be shown to be continuous and bounded; as regards the second one, it is continuous in view of Assumption 2, and the bound on its norm is affinely dependent on the norm of x. Thus we can compute constants k_{Jx}, k_{Jq0} and k_{Jq1} such that

$$\|S(q)J_x(x, z)\| \leq k_{Jx}, \quad (63a)$$

$$\|S(q)J_q(x, q)\| \leq k_{Jq0} + k_{Jq1}\,\|x\|, \quad (63b)$$

for all $q \in Q$ and $x \in R^n$.

Equation (62) results in

$$\begin{aligned}\dot W(x,z,t) \le {} & -\mu^{-1}\|z-H\|^2 \\ & + 2k_{Jx}\|z-H\|\,\|\dot x\| \\ & + 2k_{Jq0}k_q\|z-H\| \\ & + 2k_{Jq1}k_q\|z-H\|\,\|x\| \\ & + \|z-H\|^2\sum^{\eta} k_{sqi}. \end{aligned} \tag{64}$$

Since equation (51a) implies that

$$\|\dot x\| \le a_1 + a_x\|x\| + a_z\|z-H\|, \tag{65}$$

for some scalars $a_1, a_x, a_z \ge 0$, it follows that

$$\begin{aligned}\dot W(x,z,t) \le {} & -(\mu^{-1} - s_{zz})\|z-H\|^2 \\ & + s_{xz}\|x\|\,\|z-H\| + s_z\|z-h\|, \end{aligned} \tag{66}$$

with

$$s_{zz} \triangleq 2k_{Jx}a_z + \sum_{i=1}^{\eta} k_{sqi}, \tag{67a}$$

$$s_{xz} \triangleq 2(k_{Jx}a_x + k_{Jq1}k_q), \tag{67b}$$

$$s_z \triangleq 2(k_{Jx}a_1 + k_{Jq0}k_q). \tag{67c}$$

Now we are able to define the following Lyapunov function candidate.

$$L(x,z,t) = (1-\chi)V(x) + \chi W(x,z,t), \tag{68}$$

where χ is a real constant belonging to the interval (0, 1).

A bound on its derivative along the solutions of the closed loop system (51) is easily evaluated from (56) and (66) as

$$\begin{aligned}\dot L(x,z,t) \le {} & -(\|x\|\ \|z-H\|) \\ & \begin{pmatrix} (1-\chi)p_{xx} & -\dfrac{(1-\chi)p_{xz}+\chi s_{xz}}{2} \\ -\dfrac{(1-\chi)p_{xz}+\chi s_{xz}}{2} & \dfrac{\chi}{\mu}(1-\mu s_{zz}) \end{pmatrix} \\ & \begin{pmatrix} \|x\| \\ \|z-H\| \end{pmatrix} \\ & + (0\ \ \chi s_z)\begin{pmatrix} \|x\| \\ \|z-H\| \end{pmatrix} + (1-\chi)p_1. \end{aligned} \tag{69}$$

The upper bound on μ for the existence of an interval (χ_1, χ_2) of values of χ for which the first matrix of equation (69) is positive definite is given by Saberi and Khalil (1984) as

$$\mu^* = \frac{p_{xx}}{p_{xx}s_{zz} + p_{xz}s_{xz}} \tag{70}$$

Thus it is easy to verify that, for any $\mu < \mu^*$, there exists a family of functions L for which the hypotheses of the lemma reported in the Appendix apply. Uniform ultimate boundedness of the trajectories of the closed loop system (51) follows.

Remark 8. Although μ^* is actually computable via (70), it is, in general, a conservative estimate of the actual upper bound of admissible values of the perturbation parameter. Our emphasis here is on the existence of such a bound rather than its computation.

6 AN EXAMPLE

Consider the system illustrated in Fig. 1. The two bodies of masses m_1 and m_2 are connected by a massless, flexible rod characterized by a free length l_0 and a linear stiffness coefficient $k > 0$. The distance between the origins of the two coordinate systems, O_1 and O_2, is l_0; u and q are a control and an uncertain disturbance force, respectively. Moreover we assume $|q(t)| \le k_q$ and $|\dot q(t)| \le k_{\dot q}$ for some known k_q and $k_{\dot q}$.

One can recognize that, for large but finite values of k, this mechanical system has a two-time scale behavior The "slow" behavior is related to the rigid body motion of the system; the "fast" one is related to the oscillations of m_1 and m_2 at the ends of the slightly flexible bar. Since no damping is assumed in the rod, the fast behavior is only marginally stable, and this is bound to cause stability problems if one designs a controller on the basis of the slow dynamics only. Hence, our aim is to stabilize the system in the "ultimate boundedness" sense, by means of a "composite" action on the fast and slow time scales.

The motion of the system is described by

$$m_1\ddot r_1 = k(r_2 - r_1) + u, \tag{71a}$$

$$m_2\ddot r_2 = -k(r_2 - r_1) + q. \tag{71b}$$

Choosing $\mu \triangleq k^{-1/2}$, letting

$$\triangleq \begin{pmatrix} r_2 \\ \dot r_2 \end{pmatrix}, \quad z \triangleq \begin{pmatrix} k(r_2 - r_1) \\ k^{1/2}(\dot r_2 - \dot r_1) \end{pmatrix}, \tag{72}$$

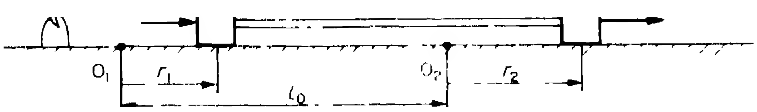


FIG. 1. A simple mechanical system

the system is described as follows:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \mu \dot{z}_1 \\ \mu \dot{z}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1/m_2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1/M_p & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1/m_1 \end{pmatrix} u + \begin{pmatrix} 0 \\ 1/m_2 \\ 0 \\ 1/m_2 \end{pmatrix} q, \quad (73)$$

with $1/M_p \triangleq 1/m_1 + 1/m_2$. System description (73) is now in the form of system (1). (For further reading on the singular perturbation approach to flexible mechanical systems, see Ficola *et al.* (1983), Marino and Nicosia (1984) and Spong (1989).)

The quasi steady state behavior of the z state, obtained by setting $\mu = 0$, is given by

$$z_{01} = -(M_p/m_1)u + (M_p/m_2)q, \quad (74a)$$

$$z_{02} = 0. \quad (74b)$$

Hence, the reduced-order system is

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1/M_t \end{pmatrix} u + \begin{pmatrix} 0 \\ 1/M_t \end{pmatrix} q, \quad (75)$$

where $M_t \triangleq m_1 + m_2$. Note that equation (75) describes the system in Fig. 1 when the rod is rigid.

It is readily seen, from equation (75) that the uncertain disturbance is matched with the input matrix.

The boundary layer system associated with equation (73) is only marginally stable. In this case one can show that, for any slow feedback law $u = -g_1 x_1 - g_2 x_2$ which stabilizes the nominal (i.e. with $q(t) \equiv 0$) reduced-order system, the full order nominal system is unstable for any $\mu > 0$.

The boundary layer system is asymptotically stabilized by any z-feedback of the form

$$u = f_1 z_1 + f_2 z_2 + \bar{u}, \quad (76)$$

with

$$f_1 > -1 - m_1/m_2, \quad f_2 > 0. \quad (77)$$

Following Spong (1989) we feedback only z_2, i.e. we let $f_1 = 0$. In this case, as can be seen using equations (74) and (76) the reduced-order system (75) is unaffected by z-feedback. Moreover, since its system matrix (A_0 in our notation) is not uncertain, the optimization described in (39)–(41) is not necessary. We have

$$k_E = 0, \quad G = 1, \quad k_L = 1. \quad (78)$$

Choosing

$$Q_0 = M_t^2 \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix},$$

the solution of Riccati equation (45) is given by

$$P_0 = M_t^2 \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}.$$

It follows that

$$v = \bar{B}_0 P_0 x = M_t(x_1 + 2x_2),$$

$$\underline{\beta}_1 = 1,$$

$$\underline{\beta}_2 = k_q.$$

The full controller, expressed in terms of the original state variables of equation (71), is as follows:

$$u = f_2 k^{1/2}(\dot{r}_2 - \dot{r}_1) - \beta_1 M_t(r_2 + 2\dot{r}_2) - \beta_2 s(\gamma M_t(r_2 + 2\dot{r}_2)), \quad (79)$$

where the function s is defined in equation (46). Theorem 4 assures us that, provided $f_2 > 0$, $\beta_1 > \underline{\beta}_1$ and $\beta_2 \geq \underline{\beta}_2$, the trajectories of the closed loop system will be uniformly ultimately bounded for k sufficiently large.

6.1. *Numerical simulation results*

Here we present some numerical simulation results carried out with $m_1 = 1$ kg, $m_2 = 2$ kg, $q(t) = 5\cos(0.5t)$ N, $k_q = 5$ N, and choosing $f_2 = 1$. With $\underline{\beta}_1 = 1$, $\underline{\beta}_2 = 5$, we set $\beta_1 = 1.5\underline{\beta}_1$, $\beta_2 = \underline{\beta}_2$ and $\gamma = 500/9$. The initial conditions were $r_1(0) = r_2(0) = 5$ m, $\dot{r}_1(0) = \dot{r}_2(0) = 0$ m sec^{-1}.

First we consider the behavior of the reduced-order system under controller (79); in Fig. 2 the time history of the displacement r_2 is plotted. The efficacy of the nonlinear component of the control law can be appreciated by comparing Fig. 2 with Fig. 3, for which the simulation has been carried out on the reduced-order model using the linear controller only, i.e. setting $\beta_2 = 0$. Figure 4 depicts the behavior of r_2 for the full order system subject to controller (79) and with $k = 10^4$ N m^{-1}. No appreciable difference can be noticed in comparing with Fig. 2. A more visible change in system behavior can be noted in Fig. 5, which corresponds to $k = 100$ N m^{-1}; however, the full order system maintains its boundedness. Figure 6 illustrates the loss of boundedness which occurs for $k = 4$ N m^{-1}.

Finally, we illustrate the consequences of not using z-feedback, i.e. letting $f_2 = 0$. With $f_2 = 0$, the response of the reduced-order system remains unchanged. However, even for $k = 10^6$ N m^{-1}, the full order system exhibits unbounded behavior; see Fig. 7.

7. CONCLUSIONS

In this paper we have considered the deterministic control of a linear singularly

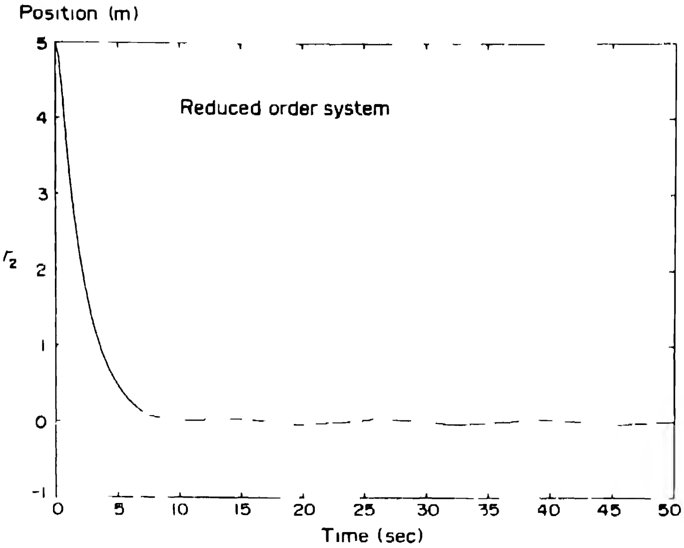


FIG 2 Time history of the position of m_2 for the reduced order model using the complete nonlinear controller

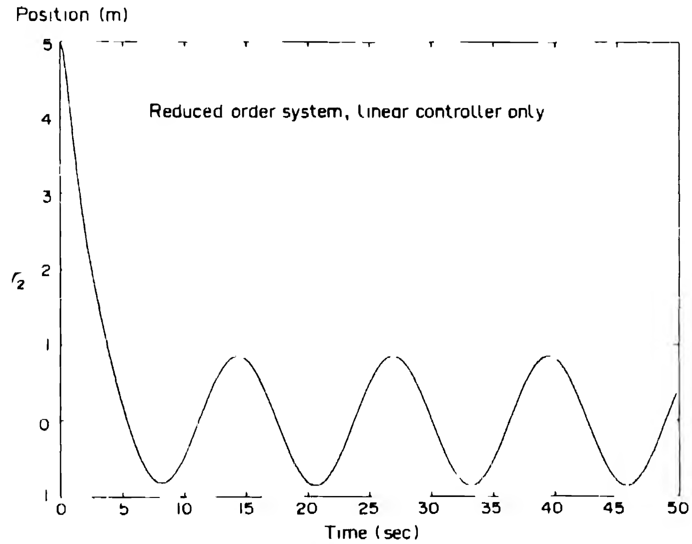


FIG 3 Time history of the position of m_2 for the reduced-order model using the linear controller only

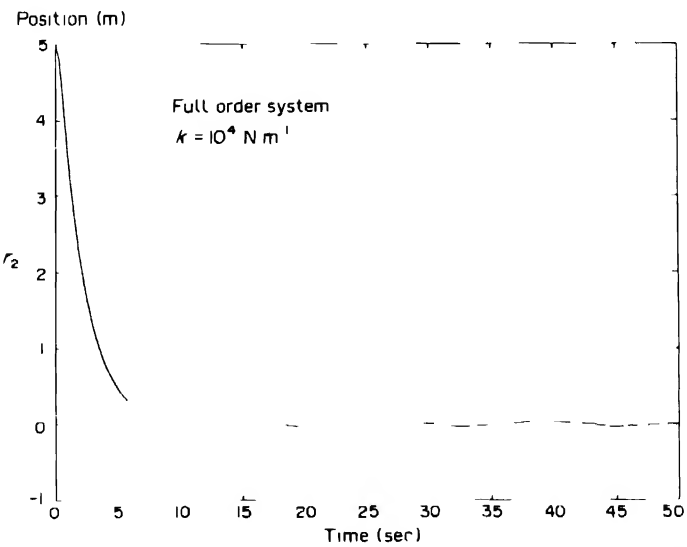


FIG 4 Time history of the position of m_2 for the full-order model with $k = 10^4$ N m $^{-1}$

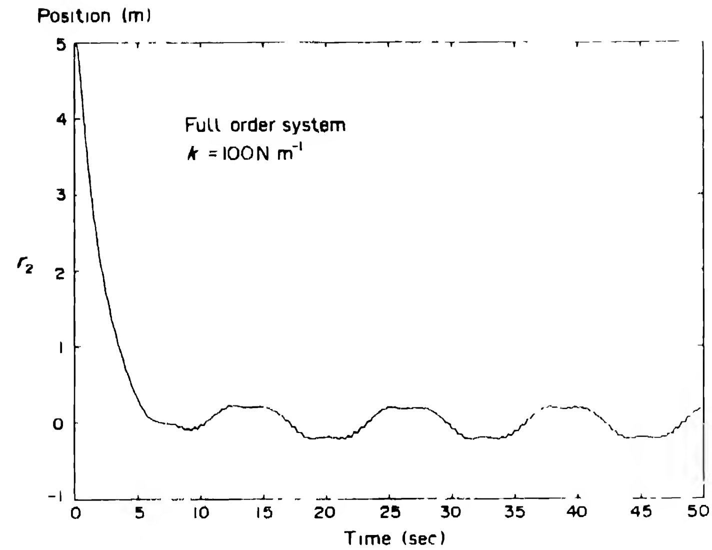


FIG. 5. Time history of the position of m_2 for the full-order model with $k = 100\,\mathrm{M\,m}^{-1}$.

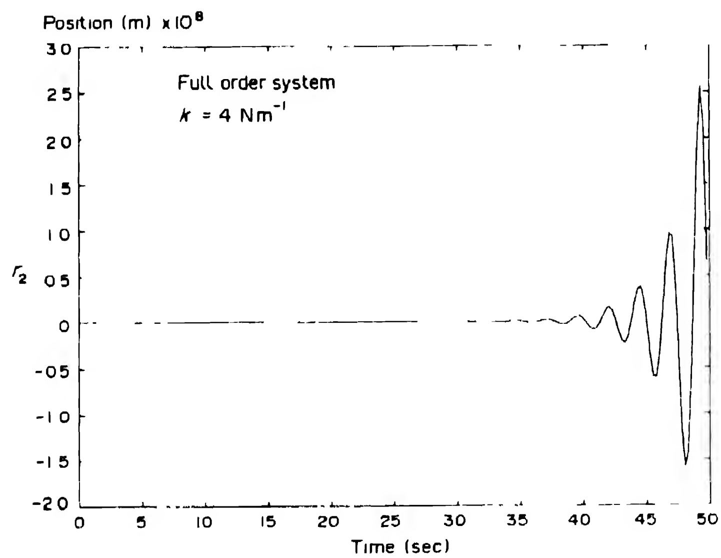


FIG. 6. Time history of the position of m_2 for the full-order model with $k = 4\,\mathrm{N\,m}^{-1}$.

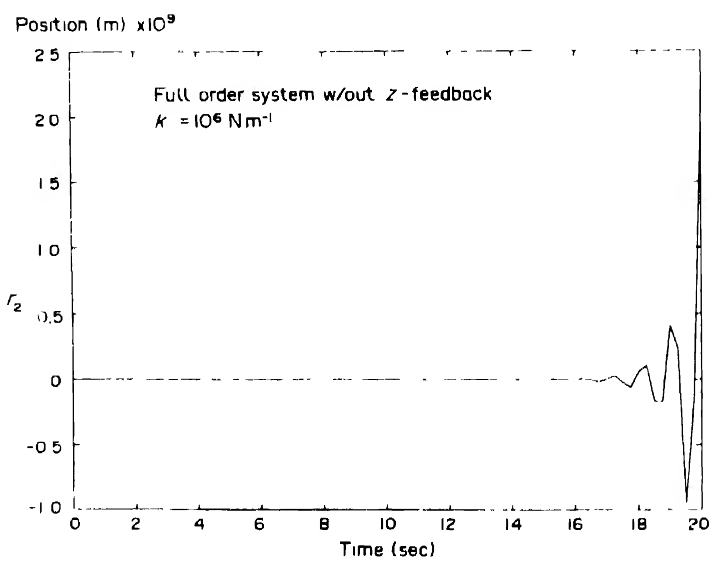


FIG. 7. Time history of the position of m_2 for the full-order model with $k = 10^6\,\mathrm{N\,m}^{-1}$ and $F = 0$. Note that the scales are different from those in the other figures.

perturbed system subject to time-varying bounded parameter uncertainties and unknown bounded disturbance inputs. We do not assume that the boundary layer system is stable. It has been shown that matching of the uncertainties with a nominal input matrix, assumed on the reduced-order system, is invariant under linear output feedback. On this basis a composite controller has been proposed, linear in the fast variable and nonlinear in the slow variable, which, for μ sufficiently small assures uniform ultimate boundedness of the trajectories of the closed loop system, independently of the uncertainties. The approach which has been followed is different from the usual composite control technique, although there are some similarities and these have been pointed out.

REFERENCES

Ambrosino, G., G. Celentano and F. Garofalo (1985). Robust model tracking control for a class of nonlinear plants. *IEEE Trans. Aut. Control*, **AC-30**, 275–279.

Calise, A. J., J. V. R. Prasad and B. Siciliano (1990). Design of optimal output feedback compensator in two-time scale systems. *IEEE Trans. Aut. Control*, **AC-35**, 488–492.

Chemouil, P. and A. M. Wahdan (1980). Output feedback control of systems with slow and fast modes. *Large Scale Sys.*, **1**, 257–264.

Corless, M. (1987). Robustness of a class of feedback-controlled uncertain nonlinear systems in the presence of singular perturbations. *Proc. Amer. Contr. Conf.*, Minneapolis, MN, pp. 1584–1589.

Corless, M. (1989). Control of uncertain mechanical systems with robustness in the presence of unmodelled flexibilities. In A. Blaquière (Ed.), *Modeling and Control of Systems in Engineering, Quantum Mechanics, Economics, and Biosciences, Lecture Notes in Control and Information Sciences*, Vol. 121, pp. 51–64. Springer-Verlag, New York.

Corless, M. and L. Glielmo (1993). Robustness of output feedback for a class of singularly perturbed nonlinear systems, (in preparation).

Corless, M. and G. Leitmann (1981). Continuous state feedback guaranteeing uniform ultimate boundedness for uncertain dynamical systems. *IEEE Trans. Aut. Control*, **AC-26**, 1139–1144.

Corless, M. and G. Leitmann (1988). Controller design for uncertain systems via Lyapunov functions. *Proc. Amer. Contr. Conf.*, pp. 2019–2025.

Corless M., F. Garofalo and L. Glielmo (1989). Nonlinear stabilization of a class of singularly perturbed uncertain systems. *IFAC Symposium on Nonlinear Control Systems Design*, Capri, Italy.

Corless, M., G. Leitmann and E. P. Ryan (1990). Control of uncertain systems with neglected dynamics. In A. S. I. Zinober (Ed.), *Deterministic Nonlinear Control of Uncertain Systems*. IEE, London.

Ficola, A., R. Marino and S. Nicosia (1983). A singular perturbation approach to the dynamic control of elastic robots. *Proc. 21st Allerton Conf. Commun. Contr. Comput.*, pp. 335–342.

Fossard, A. J. and J. F. Magni (1980). A frequential analysis of singularly perturbed systems with state or output control. *Large Scale Sys.*, **1**, 223–228.

Garofalo, F. (1988). Composite control of singularly perturbed uncertain systems with slow nonlinearities. *Int. J. Control*, **48**, 1979–1991.

Garofalo, F. and L. Glielmo (1989). Robust model tracking for a class of singularly perturbed nonlinear systems via composite control. In A. Blaquière (Ed.), *Modeling and Control of Systems in Engineering, Quantum Mechanics, Economics, and Biosciences, Lecture Notes in Control and Information Sciences*, **121**, pp. 89–110. Springer-Verlag, New York.

Garofalo, F. and G. Leitmann (1988). A composite controller ensuring ultimate boundedness for a class of singularly perturbed uncertain systems. *Dynam. Stability Sys.*, **3**, 135–145.

Garofalo, F. and G. Leitmann (1990). Nonlinear composite control of a class of nominally linear singularly perturbed uncertain systems. In A. S. I. Zinober (Ed.), *Deterministic Nonlinear Control of Uncertain Systems*. IEE, London.

Gutman, S. (1979). Uncertain dynamical systems, a Lyapunov min–max approach. *IEEE Trans. Aut. Control*, **AC-24**, 437–443.

Kailath, T. (1980). *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Kalman, R. E. and J. E. Bertram (1960). Control system analysis and design via the "second method" of Lyapunov, I: continuous-time systems. *ASME J. Basic Engineering*, **82**, 371–393.

Khalil, H. K. (1981). On the robustness of output feedback control methods to modeling errors. *IEEE Trans. Aut. Control*, **AC-26**, 524–526.

Khalil, H. K. (1984). A further note on the robustness of output feedback control methods to modeling errors. *IEEE Trans. Aut. Control*, **AC-29**, 861–862.

Khalil, H. K. (1985). Output feedback control of linear two time scale systems. *Proc. 1985 Amer. Contr. Conf.*, Boston, MA.

Kokotović, P. V. and H. K. Khalil (Eds.) (1986). *Singular Perturbations in Systems and Control*. IEEE Reprint Volume.

Kokotović, P. V., H. K. Khalil and J. O'Reilly (1986). *Singular Perturbation Methods in Control; Analysis and Design*. Academic Press, London.

Leitmann, G. (1989a). Deterministic control of uncertain systems via a constructive use of Lyapunov stability theory. *Proc. SIAM Conf. "Control in the 90s"*, San Francisco, CA.

Leitmann, G. (1989b). Controlling singularly perturbed uncertain dynamical systems. In A. Blaquière (Ed.), *Modeling and Control of Systems in Engineering, Quantum Mechanics, Economics, and Biosciences, Lecture Notes in Control and Information Sciences*, **121**, 3–14. Springer-Verlag, New York

Leitmann, G., E. P. Ryan and A. Steinberg (1986). Feedback control of uncertain systems: robustness with respect to neglected actuator and sensor dynamics. *Int. J. Control*, **43**, 1243–1256.

Marino, R. and S. Nicosia (1984). On the feedback control of industrial robots with elastic joints: a singular perturbation approach. Università degli Studi di Roma "Tor Vergata", Dipartimento di Ingegneria Elettronica, Int. Rep. R-84.01, Rome, Italy.

Moerder, D. D. and A. J. Calise (1985). Two-time scale stabilization of systems with output feedback. *J. Guidance*, **8**, 731–736.

Moerder, D. D. and A.J. Calise (1988). Near-optimal output feedback regulation of ill-conditioned linear systems. *IEEE Trans. Aut. Control*, **AC-33**, 463–466.

Ryan, E. P. and M. Corless (1991). Robust feedback control of singularly perturbed uncertain dynamical systems. *Dynam. Stability Sys.*, **6**, 107–121.

Saberi, A. and H. Khalil (1984). Quadratic-type Lyapunov functions for singularly perturbed systems. *IEEE Trans. Aut. Control*, **AC-29**, 542–550.

Spong, M. W. (1989). Adaptive control of flexible joint manipulators *Sys. Contr. Lett.*, **13**, 15–21.

Taylor, D. G., P. V. Kokotović, R. Marino and I. Kanellakopoulos (1989). Adaptive regulation of nonlinear systems with unmodelled dynamics. *IEEE Trans. Aut. Control*, **AC-34**, 405–412.

Vidyasagar, M. (1985a). Robust stabilization of singularly perturbed systems. *Sys. Contr. Lett.*, **5**, 413–418.

Vidyasagar, M. (1985b). *Control System Synthesis: A Factorization Approach*. MIT Press, Cambridge, MA.

APPENDIX

Some definitions and a lemma

We refer to a differential equation of the form

$$\dot{y}(t) = F(y(t), t), \tag{80}$$

with $y \in R^n$, $t \in R$ and $F(\cdot)$: $R^n \times R \rightarrow R^n$. We say that the trajectories of system (80) are uniformly ultimately bounded within a spherical neighbourhood of radius ρ of $y = 0$ (indicated with $B(\rho)$) iff the following properties are satisfied.

Existence of solutions. Given any $(y_0, t_0) \in R^n \times R$ there exists a solution $y(\cdot):[t_0, t_1) \rightarrow R^n$ of equation (80) with $t_1 > t_0$.

Indefinite extension of solutions. Every solution $y(\cdot):[t_0, t_1) \rightarrow R^n$ of (80) has an extension over $[t_0, \infty)$.

Global uniform boundedness. Given any bound $r \in R_+$, there exists a positive $d(r) < \infty$ such that if $y(\cdot)$ is a solution of equation (80) with $y(t_0) \in B(r)$, then $y(t) \in B(d(r))$ for all $t \geq t_0$.

Local boundedness within B(ρ). There exists a spherical neighbourhood $B(\rho_0)$ of $y = 0$ such that, if $y(\cdot)$ is a solution of equation (80) with $y(t_0) \in B(\rho_0)$ then $y(t) \in B(\rho)$ for all $t \geq t_0$.

Global uniform ultimate boundedness within B(ρ). Given any bound $r \in R_+$, there exists a $T(r) \in R_+$ such that, if $y(\cdot)$ is a solution of equation (80) with $y(t_0) \in B(r)$, then $y(t) \in B(\rho)$ for all $t \geq t_0 + T(r)$.

The listed properties of system (80) can be assured with the aid of the following lemma; see Corless and Leitmann (1981).

Lemma 1. Consider system (80) and suppose that there exist a positive $s < \infty$, a C^1 scalar function $\mathscr{L}: R^n \times R \rightarrow R_+$, class **KR** functions χ_1 and χ_2, and a class **K** function χ_3 such that

$$\chi_1(\|y\|) \leq \mathscr{L}(y, t) \leq \chi_2(\|y\|), \tag{81}$$

$$\frac{\partial \mathscr{L}}{\partial t}(t, y) + \frac{\partial \mathscr{L}}{\partial x}(y, t)F(y, t) \leq -\chi_3(\|y\|), \tag{82}$$

for all $t \in R$ and $\|y\| \geq s$. Then for any $\rho > \chi_1^{-1} \circ \chi_2(s)$, the trajectories of the system are uniformly ultimately bounded within $B(\rho)$.

Automatica Vol 29 No 2 pp 401–423 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

H^{∞}-Optimal Control for Singularly Perturbed Systems. Part I: Perfect State Measurements*†‡

ZIGANG PAN and TAMER BAŞAR§

Design of composite controllers for singularly perturbed systems under the H^{∞} criterion requires a much different methodology than that available for the singularly perturbed regulator problem

Key Words— Singular perturbations; H^{∞}-optimal control; feedback control; feedforward control; differential games; optimal regulators; boundary layer system.

Abstract—We study the H^{∞}-optimal control of singularly perturbed linear systems under perfect state measurements, for both finite and infinite horizons. Using a differential game theoretic approach, we show that as the singular perturbation parameter ϵ approaches zero, the optimal disturbance attenuation level for the full-order system under a quadratic performance index converges to a value that is bounded above by the maximum of the optimal disturbance attenuation levels for the slow and fast subsystems under appropriate slow and fast quadratic cost functions. Furthermore, we construct a composite controller based on the solution of the slow and fast games, which guarantees a desired achievable performance level for the full-order plant, as ϵ approaches zero. A slow controller, however, is not generally robust in this sense, but still, under some conditions, which are delineated in the paper, the fast dynamics can be totally ignored. The paper also studies optimality when the controller includes a feedforward term in the disturbance, and presents some numerical examples to illustrate the theoretical results.

1. INTRODUCTION

ONE OF THE IMPORTANT recent developments in control theory has been the recognition of the close relationship that exists between H^{∞}-optimal control problems and a class of linear-quadratic differential games, which has not only led to simpler derivations of existing results on the former, but also enabled us to develop worst-case (H^{∞}-optimal) controllers under various information patterns, such as (in addition to perfect and imperfect state measurements) delayed state and sampled state measurements. An up-to-date coverage of this relationship and the derivation of H^{∞}-optimal controllers under different information patterns can be found in the recent book by Başar and Bernhard (1991), which also contains an extensive list of references on the topic.

This paper makes further use of the relationship between H^{∞}-optimal control and differential games in studying the issue of worst-case controller design for systems with fast and slow dynamics—systems commonly modeled using the mathematical framework of "singular perturbations". The main objective here is to obtain 'approximate' controllers independently of the "small" singular perturbation parameter, say ϵ, and to prove that the approximate controllers can be used "reliably" on the original system when ϵ is sufficiently small. It is, of course, well-known that in the absence of the disturbance, a composite design based on two separate designs for the slow and fast subsystems performs remarkably well in the linear-quadratic (singularly perturbed) regulator problem, with the approximation to the optimal cost being $O(\epsilon^2)$ (Chow and Kokotovic, 1976). Another appealing feature of the composite controller (in the regulator problem) is that it avoids the difficulty of dealing with stiff differential equations.

One natural question here is whether the counterpart of the "singularly perturbed regulator theory" can be developed for H^{∞}-optimal control, or equivalently for the related class of linear-quadratic zero-sum differential games, by going through similar steps. One of the messages of this paper is that this is not necessarily so, and one has to develop a separate theory by going through some new lines of reasoning,

* Received 23 April 1991; revised 19 November 1991; received in final form 11 March 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor H. Kimura under the direction of Editor H. Kwakernaak.

† Research supported in part by the U.S. Department of Energy under Grant DE-FG-02-88-ER-13939, in part by the National Science Foundation under Grant ECS-1-5-30496, and in part by the Joint Services Electronics Program under Contract N00014-84-c-0149.

‡ An abridged version was presented at the 1992 American Control Conference, Chicago, IL, 24–26 June 1992.

§ Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois, 1101 W. Springfield Avenue, Urbana, IL 61801, U.S.A.

which has indeed been accomplished in the paper. To make the above statement somewhat more precise, let $\gamma^*(\epsilon)$ denote the H^∞-optimum performance of the full-order system under perfect state measurements, and γ_s and γ_f denote the H^∞-optimum performances of (appropriately defined) reduced slow and fast subsystems, respectively. Our first result is that it is not necessarily true that $\gamma^*(\epsilon = 0) = \gamma_s$, and an H^∞-norm bounding controller for the reduced-order system that guarantees a performance (for the reduced system) in the neighborhood of γ_s could lead to a catastrophic performance for the full-order system, no matter how small ϵ is. In terms of the soft-constrained game associated with the full-order system, such a feedback controller (which depends only on the slow state, by construction) may lead to an unbounded value. This is quite in contrast with the limiting properties of feedback controllers in disturbance-free optimal control problems with singularly perturbed dynamics, but is not that surprising in view of some existing results on singularly perturbed Nash games (Gardner and Cruz, 1978) where it is known that the well-posedness (or, continuity at $\epsilon = 0$) depends on the consistency of the information structure between the full and reduced-order problems. Zero-sum differential games fall somewhere between optimal control problems and nonzero-sum differential games, and there the saddle-point value shows continuity whenever it exists for both the full and reduced-order problems, while the condition of existence depends on the consistency of the underlying information structures. But this is precisely what determines the optimal level of performance and the corresponding minimax controller in H^∞-optimization, which therefore requires a careful study of the dependence of conjugate points in zero-sum differential games on the small (singular perturbation) parameter ϵ. This paper undertakes such a study and provides a complete analysis of the singularly perturbed H^∞-optimal control problem under perfect state measurements. It is proven that $\gamma^*(\epsilon = 0) \leq \max\{\gamma_s, \gamma_f\}$, and that a composite controller exists under which such a performance bound is attained "approximately".

Design of controllers for singularly perturbed systems subject to unknown disturbances has been studied before in the literature, notably in papers by Garofalo and Leitmann (1988, 1990) and Corless *et al.* (1990), where the objective has been to obtain composite controllers that guarantee stability of the overall (possibly nonlinear) system. The main approach of the authors in these papers has involved the construction of appropriate Liapunov functions, in terms of which a class of stabilizing controllers has been characterized. However, no optimality properties have been associated with these controllers, which is our main concern in this paper. Yet another paper that deals with uncertain (linear) systems which exhibit time-scale separation is the one by Luse and Ball (1989), which obtains a two-frequency scale decomposition for H^∞-disk problems, but does not address the issue of optimal controller design.

The present paper is organized as follows. In the next section (Section 2) we formulate the singularly perturbed H^∞-optimal control problem and the associated linear-quadratic differential game, both under perfect state measurements. We also provide in that section the solution to the full-order problem, for both finite and infinite horizons. In Section 3, we identify the slow and fast subsystems and the associated differential games. There are in fact two "slow" games, differing only in the information available to the controller, both of which are relevant to the problem at hand. The saddle-point solutions to these "slow" games as well as to the "fast" game are also presented in Section 3. In Section 4, where the main results are presented, we obtain composite controllers in finite and infinite horizons, prove optimality properties of these composite controllers in the context of the original H^∞-optimal control problem and obtain precise (tight) performance bounds attained by them. Section 5 presents some numerical results to illustrate the theory, and Section 6 provides a discussion on extensions to other information patterns; some of this work has already been completed and will appear elsewhere, and some is currently underway. The paper ends with four Appendices, which provide the details of some of the derivations given in the main body of the paper.

2. PROBLEM FORMULATION

The system under consideration, with slow and fast dynamics, is described in the standard "singularly perturbed" form by

$$\begin{aligned} \dot{x}_1 &= A_{11}(t)x_1 + A_{12}(t)x_2 \\ &\quad + B_1(t)u + D_1(t)w; \quad x_1(0) = 0, \\ \epsilon\dot{x}_2 &= A_{21}(t)x_1 + A_{22}(t)x_1 \\ &\quad + B_2(t)u + D_2(t)w; \quad x_2(0) = 0, \end{aligned} \tag{2.1}$$

where $x' := (x_1', x_2')$ is the n-dimensional state vector, with x_1 of dimension n_1 and x_2 of dimension $n_2 := n - n_1$; u is the control input, and w is the disturbance, each belonging to

appropriate ($\mathscr{L}^2$) Hilbert spaces $\mathscr{H}_x$, $\mathscr{H}_u$ and $\mathscr{H}_w$, respectively, defined on the time interval $[0, t_f]$. The control input u is generated by a closed-loop control policy μ, according to

$$u(t) = \mu(t, x_{[0,t]}), \tag{2.2}$$

where $\mu : [0, t_f] \times \mathscr{H}_x \to \mathscr{H}_u$ is piecewise continuous in t and Lipschitz continuous in x, further satisfying the given causality condition. Let us denote class of all these controllers by $\mathscr{M}$. With this system, we associate the standard quadratic performance index:

$$L(u, w) = |x(t_f)|^2_{Q_f} + \int_0^{t_f} (|x(t)|^2_{Q(t)} + |u(t)|^2)\, dt$$
$$\equiv |x(t_f)|^2_{Q_f} + \|x\|^2_Q + \|u\|^2; \quad Q_f \geq 0, \quad Q(\cdot) \geq 0, \tag{2.3}$$

where Q_f will show dependence on $\epsilon > 0$, as to be clarified later. Let us also introduce the notation $J(\mu, w)$ to denote $L(u, w)$, with u given by (2.2). The H^∞-optimal control problem is the minimization of the quantity

$$\sup \{J(\mu, w)\}^{1/2}/\|w\| \tag{2.4}$$

over all permissible controllers μ, and in the case a minimum does not exist, the derivation of a controller μ that will assure a performance within a given neighborhood of the infimum of (2.4). Let us denote this infimum by $\gamma^*(\epsilon)$, i.e.

$$\inf_{\mu \in \mathscr{M}} \sup_{w \in \mathscr{H}_w} \{J(\mu, w)\}^{1/2}/\|w\| = \gamma^*(\epsilon), \tag{2.5}$$

where we explicitly show the dependence of γ^* on the singular perturbation parameter $\epsilon > 0$.

For each $\epsilon > 0$, we can associate a soft-constrained linear-quadratic differential game with this worst-case design problem (see Başar and Bernhard, 1991) which has the cost function

$$L_\gamma(u, w) = L(u, w) - \gamma^2 \|w\|^2. \tag{2.6}$$

The performance level $\gamma^*(\epsilon)$ in (2.5) is then the "smallest" value of $\gamma \geq 0$ under which the differential game with state equation (2.1) and cost function (2.6) has a bounded upper value, when u is chosen according to (2.2). We also know that for every fixed $\epsilon > 0$, and for each $\gamma > \gamma^*(\epsilon)$, this differential game admits a saddle-point solution, with the saddle-point controller μ^* being a linear feedback law (see Başar and Bernhard, 1991).

Even though the problem formulated above has been solved completely for every $\epsilon > 0$, the computation of $\gamma^*(\epsilon)$ and that of a corresponding H^∞-optimal or suboptimal controller for small values of $\epsilon > 0$ present serious difficulties, due to numerical stiffness. To remedy this, we pose in this paper the question of whether $\gamma^*(\epsilon)$ and the H^∞-optimal controller can be determined, for small values of $\epsilon > 0$, by solving well-behaved ϵ-independent smaller-order problems, as in the case of the singularly perturbed linear-quadratic regulator problem (Chow and Kokotović, 1976). Another question of interest is whether the fast subsystem dynamics can be neglected completely in the design of such controllers, and if so under what conditions.

Before studying these questions, we first present below the solution to the full-order problem, under three basic assumptions:

A1. Q_f and $Q(\cdot)$ in (2.3) are partitioned as

$$Q_f = \begin{bmatrix} Q_{f11} & \epsilon Q_{f12} \\ \epsilon Q_{f21} & \epsilon Q_{f22} \end{bmatrix}; \quad Q(t) = \begin{bmatrix} Q_{11}(t) & Q_{12}(t) \\ Q_{21}(t) & Q_{22}(t) \end{bmatrix},$$

where in each case the 11-block is of dimension $n_1 \times n_1$, and the 22-block is of dimension $n_2 \times n_2$.

A2. The matrix functions $A_{ij}(t)$, $Q_{ij}(t)$, $B_i(t)$, $D_i(t)$ $(i = 1, 2, j = 1, 2)$ are continuously differentiable in $t \geq 0$.

A3. The matrices $A_{22}(t)$ and $Q_{22}(t)$ are invertible for all $t \in [0, t_f]$.

Let us further introduce the notation $A(t)$ and $A_\epsilon(t)$, to denote the partitioned matrices

$$A(t) := \begin{bmatrix} A_{11}(t) & A_{12}(t) \\ A_{21}(t) & A_{22}(t) \end{bmatrix},$$

$$A_\epsilon(t) := \begin{bmatrix} A_{11}t & A_{12}(t) \\ (1/\epsilon)A_{21}(t) & (1/\epsilon)A_{22}(t) \end{bmatrix}, \tag{2.7}$$

where we take $\epsilon > 0$. Similarly, we introduce the partitioned matrices

$$B(t) := \begin{bmatrix} B_1(t) \\ B_2(t) \end{bmatrix}; \quad B_\epsilon(t) := \begin{bmatrix} B_1(t) \\ (1/\epsilon)B_2(t) \end{bmatrix}, \tag{2.8}$$

$$D(t) := \begin{bmatrix} D_1(t) \\ D_2(t) \end{bmatrix}; \quad D_\epsilon(t) := \begin{bmatrix} D_1(t) \\ (1/\epsilon)D_2(t) \end{bmatrix}, \tag{2.9}$$

and define

$$S(t; \gamma) := B(t)B'(t) - \frac{1}{\gamma^2} D(t)D'(t),$$
$$S_\epsilon(t; \gamma) := B_\epsilon(t)B'_\epsilon(t) - \frac{1}{\gamma^2} D_\epsilon(t)D'_\epsilon(t), \tag{2.10}$$

with the ijth block of $S(t; \gamma)$ denoted henceforth by $S_{ij}(t; \gamma)$, $i,j = 1,2$.

We know from the existing theory on linear-quadratic differential games (see Başar and Olsder, 1982; Başar and Bernhard, 1991) that for each $\epsilon > 0$, there exists a $\hat{\gamma}(\epsilon) \geq 0$ such that for all $\gamma > \hat{\gamma}(\epsilon)$ the zero-sum differential

game described by (2.1) and (2.6) has a bounded upper value (which in this case is equal to zero, because the initial states are zero) as well as a saddle point, with a saddle-point controller being the feedback law

$$u^*(t) = \mu^*(t, x(t)) = -B'_\epsilon \bar{Z}(t;\epsilon)x(t), \quad t \geq 0, \tag{2.11}$$

where $\bar{Z}(t;\epsilon)$ is the unique bounded nonnegative definite solution of the generalized matrix Riccati differential equation (GRDE)

$$\dot{\bar{Z}} + A'_\epsilon \bar{Z} + \bar{Z}A_\epsilon - \bar{Z}S_\epsilon \bar{Z} + Q = 0; \quad \bar{Z}(t_f) = Q_f. \tag{2.12}$$

For $\gamma < \hat{\gamma}(\epsilon)$, on the other hand, this GRDE has a conjugate point in the open interval $(0, t_f)$, which says that it has finite escape, and furthermore the associated soft-constrained game has unbounded upper value. The level $\hat{\gamma}(\epsilon)$ is indeed the H^∞-optimal performance $\gamma^*(\epsilon)$ given by (2.5).

Now, to study the behavior of the solution of (2.12) for small values of $\epsilon > 0$, we first partition $\bar{Z}$ as follows (in a way quite analogous to the partitioning in the standard regulator problem, and consistent with the given partitioning on Q_f):

$$\bar{Z} := \begin{bmatrix} \bar{Z}_{11} & \epsilon\bar{Z}_{12} \\ \epsilon\bar{Z}_{21} & \epsilon\bar{Z}_{22} \end{bmatrix}, \tag{2.13}$$

where $\bar{Z}_{12} = \bar{Z}'_{21}$. Then, by substitution of this structure into (2.12), it is a simple matter to see that $\bar{Z}_{11}$, $\bar{Z}_{12}$ and $\bar{Z}_{22}$ satisfy the following matrix differential equations:

$$\dot{\bar{Z}}_{11} + A'_{11}\bar{Z}_{11} + A'_{21}\bar{Z}'_{12} + \bar{Z}_{11}A_{11} + \bar{Z}_{12}A_{21} + Q_{11} - \bar{Z}_{11}S_{11}\bar{Z}_{11} - \bar{Z}_{12}S_{21}\bar{Z}_{11} - \bar{Z}_{11}S_{12}\bar{Z}'_{12} - \bar{Z}_{12}S_{22}\bar{Z}'_{12} = 0; \quad \bar{Z}_{11}(t_f) = Q_{f11}, \tag{2.14}$$

$$\epsilon\dot{\bar{Z}}_{12} + \epsilon A'_{11}\bar{Z}_{12} + A'_{21}\bar{Z}_{22} + \bar{Z}_{11}A_{12} + \bar{Z}_{12}A_{22} + Q_{12} - \epsilon\bar{Z}_{11}S_{11}\bar{Z}_{12} - \epsilon\bar{Z}_{12}S_{21}\bar{Z}_{12} - \bar{Z}_{11}S_{12}\bar{Z}_{22} - \bar{Z}_{12}S_{22}\bar{Z}_{22} = 0; \quad \bar{Z}_{12}(t_f) = Q_{f12}, \tag{2.15}$$

$$\epsilon\dot{\bar{Z}}_{22} + \epsilon A'_{12}\bar{Z}_{12} + A'_{22}\bar{Z}_{22} + \epsilon\bar{Z}'_{12}A_{12} + \bar{Z}_{22}A_{22} + Q_{22} - \epsilon^2\bar{Z}'_{12}S_{11}\bar{Z}_{12} - \epsilon\bar{Z}_{22}S_{21}\bar{Z}_{12} - \epsilon\bar{Z}'_{12}S_{12}\bar{Z}_{22} - \bar{Z}_{22}S_{22}\bar{Z}_{22} = 0; \quad \bar{Z}_{22}(t_f) = Q_{f22}. \tag{2.16}$$

Let Z_{11}, Z_{12} and Z_{22} denote the limiting solutions to above equations in the open interval $(0, t_f)$, as $\epsilon \downarrow 0$†. Then Z_{11}, Z_{12} and Z_{22} are given by the following differential and algebraic equations, which are obtained by setting $\epsilon = 0$ in (2.14)–(2.16):

$$\dot{Z}_{11} + A'_{11}Z_{11} + A'_{21}Z'_{12} + Z_{11}Z_{11} + Z_{12}A_{21} + Q_{11} - Z_{11}S_{11}Z_{11} - Z_{12}S_{21}Z_{11} - Z_{11}S_{12}Z'_{12} - Z_{12}S_{22}Z'_{12} = 0; \quad Z_{11}(t_f) = Q_{f11}, \tag{2.17}$$

$$A'_{21}Z_{22} + Z_{11}A_{12} + Z_{12}A_{22} + Q_{12} - Z_{11}S_{12}Z_{22} - Z_{12}S_{22}Z_{22} = 0, \tag{2.18}$$

$$A'_{22}Z_{22} + Z_{22}A_{22} + Q_{22} - Z_{22}S_{22}Z_{22} = 0. \tag{2.19}$$

Since as $t \to t_f$, the solution Z_{12} and Z_{22} above will not, in general, satisfy the given terminal conditions (in (2.15) and (2.16), respectively) we have to introduce a boundary layer correction term (see Kokotović and Yackel, 1972). Toward this end, introduce the fast time variable $\tau = (t - t_f/\epsilon)$, and rewrite (2.15) and (2.16) on this time scale, as $\epsilon \downarrow 0$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}L_{12}(\tau) = -L_{12}(\tau)(A_{22}(t_f) - S_{22}(t_f)L_{22}(\tau)) - (A'_{21}(t_f) - Z_{11}(t_f)S_{12}(t_f))L_{22}(\tau) - Z_{11}(t_f)A_{12}(t_f) - Q_{12}(t_f); \quad L_{12}(0) = Q_{f12}, \tag{2.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}L_{22}(\tau) = -A'_{22}(t_f)L_{22}(\tau) - L_{22}(\tau)A_{22}(t_f) - Q_{22}(t_f) + L_{22}(\tau)S_{22}(t_f)L_{22}(\tau); \quad L_{22}(0) = Q_{f22}. \tag{2.21}$$

Define

$$Z'_{\mathrm{fb}}(\tau) := L_{12}(\tau) - Z_{12}(t_f), \tag{2.22}$$

$$Z_{\mathrm{fb}}(\tau) := L_{22}(\tau) - Z_{22}(t_f). \tag{2.23}$$

Then the solution of (2.14)–(2.16) will be approximated by

$$\bar{Z}_{11}(t) = Z_{11}(t) + Q(\epsilon),$$

$$\bar{Z}_{12}(t) = Z_{12}(t) + Z'_{\mathrm{fb}}\left(\frac{t - t_f}{\epsilon}\right) + O(\epsilon), \quad \forall t \in [0, t_f], \tag{2.24}$$

$$\bar{Z}_{22}(t) = Z_{22}(t) + Z_{\mathrm{fb}}\left(\frac{t - t_f}{\epsilon}\right) + O(\epsilon).$$

Conditions under which this approximation is valid will be delineated later in Theorem 2 of Section 4.

Of course, the preceding analysis is valid provided that γ is larger than $\sup\{\hat{\gamma}(\epsilon), 0 < \epsilon \leq \epsilon_0\}$, where ϵ_0 is some prechosen small (positive) scalar.

To study the infinite horizon case (i.e. as $t_f \to \infty$, as well as when $t_f = \infty$), we take A, B, D, Q to be time-invariant, and $Q_f = 0$. Furthermore, we assume that (A_ϵ, B_ϵ) is controllable, and (A_ϵ, Q) is observable for every $\epsilon > 0$. Then,

† The existence of such limits will be shown later in this section (see Theorem 2)

for each $\epsilon > 0$, there exists a $\hat{\gamma}_\infty(\epsilon)$ such that, for all $\gamma > \hat{\gamma}_\infty(\epsilon)$, the infinite-horizon soft-constrained game has a finite upper value, achieved by the time-invariant feedback controller

$$u^*(t) = \mu^\infty(x) = -B_\epsilon'\bar{Z}_\infty(\epsilon)x(t), \quad t \geq 0, \quad (2.25)$$

where $\bar{Z}_\infty > 0$ is the minimal positive definite solution of the generalized algebraic Riccati equation (GARE)

$$A_\epsilon'\bar{Z}_\infty + \bar{Z}_\infty A_\epsilon - \bar{Z}_\infty S_\epsilon \bar{Z}_\infty + Q = 0. \quad (2.26)$$

The level $\hat{\gamma}_\infty(\epsilon)$ is again the H^∞-optimal performance level ($\gamma_\infty^*(\epsilon)$) of the infinite-horizon disturbance attenuation problem, and for $\gamma < \hat{\gamma}_\infty(\epsilon)$, the soft-constrained game has infinite upper value.

As in the finite-horizon case, we now substitute the structure (2.13) into (2.26), to arrive at the following coupled algebraic equations for $\bar{Z}_{\infty 11}$, $\bar{Z}_{\infty 12}$ and $\bar{Z}_{\infty 22}$:

$$\begin{aligned} &A_{11}'\bar{Z}_{\infty 11} + A_{21}'\bar{Z}_{\infty 12}' + \bar{Z}_{\infty 11}A_{11} + \bar{Z}_{\infty 12}A_{21} \\ &\quad + Q_{11} - \bar{Z}_{\infty 11}S_{11}\bar{Z}_{\infty 11} - \bar{Z}_{\infty 12}S_{21}\bar{Z}_{\infty 11} \\ &\quad - \bar{Z}_{\infty 11}S_{12}\bar{Z}_{\infty 12}' - \bar{Z}_{\infty 12}S_{22}\bar{Z}_{\infty 12}' = 0, \quad (2.27) \end{aligned}$$

$$\begin{aligned} &\epsilon A_{11}\bar{Z}_{\infty 12} + A_{21}'\bar{Z}_{\infty 22} + \bar{Z}_{\infty 11}A_{12} + \bar{Z}_{\infty 12}A_{22} \\ &\quad + Q_{12} - \epsilon\bar{Z}_{\infty 11}S_{11}\bar{Z}_{\infty 12} - \epsilon\bar{Z}_{\infty 12}S_{21}\bar{Z}_{\infty 12} \\ &\quad - \bar{Z}_{\infty 11}S_{12}\bar{Z}_{\infty 22} - \bar{Z}_{\infty 12}S_{22}\bar{Z}_{\infty 22} = 0, \quad (2.28) \end{aligned}$$

$$\begin{aligned} &\epsilon A_{12}'\bar{Z}_{\infty 12} + A_{22}'\bar{Z}_{\infty 22} + \epsilon\bar{Z}_{\infty 12}'A_{12} + \bar{Z}_{\infty 22}A_{22} \\ &\quad + Q_{22} - \epsilon^2\bar{Z}_{\infty 12}'S_{11}\bar{Z}_{\infty 12} - \epsilon\bar{Z}_{\infty 22}S_{21}\bar{Z}_{\infty 12} \\ &\quad - \epsilon\bar{Z}_{\infty 12}'S_{12}\bar{Z}_{\infty 22} - \bar{Z}_{\infty 22}S_{22}\bar{Z}_{\infty 22} = 0. \quad (2.29) \end{aligned}$$

Letting $Z_{\infty 11}$, $Z_{\infty 12}$ and $Z_{\infty 22}$ denote the limiting solutions to the above equations as $\epsilon \downarrow 0$,† we obtain (by setting $\epsilon = 0$ in (2.27)–(2.29)) that $Z_{\infty 11}$, $Z_{\infty 12}$ and $Z_{\infty 22}$ satisfy the following coupled matrix equations:

$$\begin{aligned} &A_{11}'Z_{\infty 11} + A_{21}'Z_{\infty 21} + Z_{\infty 11}A_{11} + Z_{\infty 12}A_{21} + Q_{11} \\ &\quad - Z_{\infty 11}S_{11}Z_{\infty 11} - Z_{\infty 12}S_{21}Z_{\infty 11} - Z_{\infty 11}S_{12}Z_{\infty 21} \\ &\quad - Z_{\infty 12}S_{22}Z_{\infty 21} = 0, \quad (2.30) \end{aligned}$$

$$\begin{aligned} &A_{21}'Z_{\infty 22} + Z_{\infty 11}A_{12} + Z_{\infty 12}A_{22} + Q_{12} \\ &\quad - Z_{\infty 11}S_{12}Z_{\infty 22} - Z_{\infty 12}S_{22}Z_{\infty 22} = 0, \quad (2.31) \end{aligned}$$

$$A_{22}'Z_{\infty 22} + Z_{\infty 22}A_{22} + Q_{22} - Z_{\infty 22}S_{22}Z_{\infty 22} = 0. \quad (2.32)$$

Thus completing the analysis of the direct solution to the full-order problem as $\epsilon \downarrow 0$, we now turn, in the next two sections, to the original goal of this paper—which is the derivation of the approximate solution based on a time-scale decomposition. First we identify, in Section 3, the slow and fast subsystems associated with the original problem, and obtain the solutions of two separate H^∞-optimal control problems, one defined on the slow time scale and the other one on the fast time scale.

† The existence of such limits will be verified later in Section 4 (see Theorem 2).

3. A TIME-SCALE DECOMPOSITION

3.1. *The slow subsystem and the associated soft-constrained game*

To obtain the slow dynamics associated with (2.1), we let $\epsilon = 0$ and solve for x_2 (to be denoted $\bar{x}_2$) in terms of $x_1 =: x_s$, $u =: u_s$, $w =: w_s$, and under the working Assumption A3:

$$\bar{x}_2 = -A_{22}^{-1}(A_{21}x_s + B_2u_s + D_2w_s). \quad (3.1)$$

Using this in the first equation of (2.1), we obtain the reduced-order (slow) dynamics:

$$\dot{x}_s = A_0x_s + B_0u_s + D_0w_s, \quad (3.2)$$

where

$$A_0 := A_{11} - A_{12}A_{22}^{-1}A_{21}, \quad (3.3)$$

$$B_0 := B_1 - A_{12}A_{22}^{-1}B_2, \quad (3.4)$$

$$D_0 := D_1 - A_{12}A_{22}^{-1}D_2. \quad (3.5)$$

Using (3.1) also in the cost function (2.3) leads to the reduced (slow) cost (with $x_1 = x_s$):

$$\begin{aligned} L_s = |x_s(t_f)|^2_{Q_{f11}} + \int_0^{t_f} (&|x_s|^2_{Q_{11}} + x_s'Q_{12}\bar{x}_2 \\ &+ \bar{x}_2'Q_{21}x_s + |\bar{x}_2|^2_{Q_{22}} + |u_s|^2)\,dt. \quad (3.6) \end{aligned}$$

In view of our earlier discussion, the H^∞-optimal control problem with state equation (3.2) and cost function (3.6) is closely related to the zero-sum differential game with the same state equation and with cost function

$$L_{\gamma s} = L_s - \gamma^2\|w_s\| \quad (3.7)$$

and the quantity of interest is the upper value of this game. But note that this differential game (henceforth called the "slow game") is structurally somewhat different from the "full" game discussed in the previous section, because the cost function (3.6) now includes "cross terms" between the state x_s, control u_s, and disturbance w_s. The presence of these cross terms makes it necessary to distinguish between two different types of information structures: one where the control is allowed to depend only on the state (and not directly on the disturbance), and another where the control is allowed to depend on both (the current values of) the state and the disturbance. We will refer to a controller of the former type as state feedback (SF) controller, and to one of the latter type as full information

(FI) controller. As we shall see shortly, these controllers require different existence conditions, and hence lead to different optimal values for γ—of course the one associated with the SF controller being more restrictive than the one obtained under the FI controller. In the absence of the cross terms, however, no such distinction exists (see Başar and Bernhard, 1991).

We now study the solution to the slow differential game under these two controllers (or information structures) separately, since both will be needed in the analysis to follow in Section 4. Let us first consider the case of the FI controller, which essentially involves a disturbance feedforward term. Compatible with the given information structure, we introduce the transformations

$$\begin{aligned}\bar{u}_s &= (I + B_2'A_{22}'^{-1}Q_{22}A_{22}^{-1}B_2)^{1/2}\\ &\quad\cdot[u_s + (I + B_2'A_{22}'^{-1}Q_{22}A_{22}^{-1}B_2)^{-1}\\ &\quad\cdot B_2'A_{22}'^{-1}Q_{22}A_{22}^{-1}((A_{21} - A_{22}Q_{22}^{-1}Q_{21})x_s + D_2w_s)],\end{aligned} \tag{3.8}$$

$$\begin{aligned}\bar{w}_s &= (\gamma^2 I - D_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}D_2)^{1/2}\\ &\quad\cdot[w_s + (\gamma^2 - D_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}D_2)^{-1}\\ &\quad\cdot[D_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}\\ &\quad\cdot(A_{22}Q_{22}^{-1}Q_{21} - A_{21}x_s)],\end{aligned} \tag{3.9}$$

where the latter one is valid under the condition

$$\gamma^2 I - D_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}D_2 > 0, \tag{3.10}$$

which can equivalently be written as

$$S_{22}(\gamma) + A_{22}Q_{22}^{-1}A_{22}' > 0, \tag{3.11}$$

where

$$S_{22}(\gamma) := B_2B_2' - \frac{1}{\gamma^2}D_2D_2'. \tag{3.12}$$

If we use (3.8) and (3.9) in (3.2) and (3.7), we arrive at the following standard LQ differential game, which has no cross terms between state, control and disturbance in the cost (see Appendix A for details of the underlying manipulations):

$$\dot{x}_s = \bar{A}_0x_s + \bar{B}_0\bar{u}_s + \bar{D}_0\bar{w}_s;\quad x_s(0) = 0, \tag{3.13}$$

$$L_{\gamma s} = |x_s(t_f)|^2_{Q_{f1}} + \int_0^{t_f}(|x_s|^2_{\bar{Q}} + |\bar{u}_s|^2 - |\bar{w}_s|^2)\,dt. \tag{3.14}$$

The coefficient matrices above are explicit functions of the parameter γ, and are written as:

$$\begin{aligned}\bar{A}_0(\gamma) &= A_{11} - A_{12}Q_{22}^{-1}Q_{21} - (S_{12} + A_{12}Q_{22}^{-1}A_{22}')\\ &\quad\cdot(S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}\cdot(A_{21} - A_{22}Q_{22}^{-1}Q_{21}),\end{aligned} \tag{3.15}$$

$$\bar{B}_0(\gamma) = B_0(I + B_2'A_{22}'^{-1}Q_{22}A_{22}^{-1}B_2)^{-1/2}, \tag{3.16}$$

$$\begin{aligned}\bar{D}_0(\gamma) &= (D_0 - B_0B_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}D_2)\\ &\quad\cdot(\gamma^2 I - D_2'(A_{22}Q_{22}^{-1}A_{22}' + B_2B_2')^{-1}D_2)^{-1/2},\end{aligned} \tag{3.17}$$

$$\begin{aligned}\bar{Q}(\gamma) &= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (A_{21}' - Q_{12}Q_{22}^{-1}A_{22}')\\ &\quad\cdot S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}\cdot(A_{21} - A_{22}Q_{22}^{-1}Q_{21}).\end{aligned} \tag{3.18}$$

For each $\gamma \ge 0$, we associate the following generalized RDE with this standard LQ differential game:

$$\begin{aligned}\dot{\bar{Z}}_s + \bar{A}_0'\bar{Z}_s + \bar{Z}_s\bar{A}_0 - \bar{Z}_s(\bar{B}_0\bar{B}_0'\\ - \bar{D}_0\bar{D}_0')\bar{Z}_s + \bar{Q} = 0;\quad \bar{Z}_s(t_f) = Q_{f11}.\end{aligned} \tag{3.19}$$

Let us introduce

$$S_0 := \bar{B}_0\bar{B}_0' - \bar{D}_0\bar{D}_0', \tag{3.20}$$

which can be rewritten in terms of the original system matrices as follows (see Appendix A for details):

$$\begin{aligned}S_0 &= S_{11} + A_{12}Q_{22}^{-1}A_{12}'\\ &\quad - (S_{12} + A_{12}Q_{22}^{-1}A_{22}')(S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}\\ &\quad\cdot(S_{21} + A_{22}Q_{22}^{-1}A_{12}').\end{aligned} \tag{3.21}$$

In view of this, let us introduce the set:

$$\begin{aligned}\Gamma_s := \{\gamma' > 0 : \forall\gamma \ge \gamma', S_{22}(\gamma) + A_{22}Q_{22}^{-1}A_{22}' > 0\\ \text{and (3.19) has a bounded nonnegative}\\ \text{definite solution over } [0, t_f].\}\end{aligned} \tag{3.22}$$

and further define

$$\gamma_s := \inf\{\gamma \in \Gamma_s\}. \tag{3.23}$$

Then, it follows from the analysis of Başar and Bernhard (1991), Chapter 8, that the transformed game with cost function $L_{\gamma s}$ has a bounded upper value† if $\gamma > \gamma_s$, and only if $\gamma \ge \gamma_s$. For $\gamma > \gamma_s$, let $Z_{s\gamma}$ be the unique nonnegative definite solution of (3.19). Then, there exist strongly time consistent (feedback) saddle-point policies for the transformed game, given by

$$\bar{u}_{s\gamma}^* = \bar{\mu}_{s\gamma}^*(t, x_s(t)) = -\bar{B}_0'Z_{s\gamma}x_s(t), \tag{3.24}$$

$$\bar{w}_{s\gamma}^* = \bar{v}_{s\gamma}^*(t, x_s(t)) = \bar{D}_0'Z_{s\gamma}x_s(t). \tag{3.25}$$

Applying the inverse transformation of (3.9) to (3.25), we arrive (after some matrix manipulations, details of which can be found in Appendix A) at the following expression:

$$\begin{aligned}w_{s\gamma}^* = v_{s\gamma}^*(t, x_s(t)) &= \frac{1}{\gamma^2}(D_1'Z_{s\gamma} - D_2'(S_{22}\\ &\quad + A_{22}Q_{22}^{-1}A_{22}')^{-1}((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma}\\ &\quad - (A_{21} - A_{22}Q_{22}^{-1}Q_{21})))x_s(t).\end{aligned} \tag{3.26}$$

† Note that under condition (3.11), $\bar{Q}(\gamma) \ge 0$.

The inverse transformation of (3.8) applied to (3.24) yields:

$$\begin{aligned} u^*_{sf\gamma} &= \mu^*_{sf\gamma}(t, x_s(t), w_s(t)) \\ &= -(I + B_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} B_2)^{-1} \\ &\quad \cdot (B_0' Z_{s\gamma} - B_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} \\ &\quad \cdot (A_{22} Q_{22}^{-1} Q_{21} - A_{21})) x_s(t) \\ &\qquad - B_2'(B_2 B_2' + A_{22} Q_{22}^{-1} A_{22}')^{-1} D_2 w_s, \quad (3.27) \end{aligned}$$

which is a controller that achieves the finite upper value (which is zero) in the slow FI game Note that this controller involves a feedforward term in w_s, as expected. If, however, we wish to obtain one which depends only on the state, we simply substitute (3.26) into (3.27), to arrive at the following expression:

$$\begin{aligned} u^*_{s\gamma} &= \mu^*_{s\gamma}(t, x_s(t)) \\ &= (-B_1' Z_{s\gamma} + B_2'(S_{22} + A_{22} Q_{22}^{-1} A_{22}')^{-1} \\ &\quad \cdot ((S_{21} + A_{22} Q_{22}^{-1} A_{12}') Z_{s\gamma} \\ &\qquad - (A_{21} - A_{22} Q_{22}^{-1} Q_{21}))) x_s(t). \quad (3.28) \end{aligned}$$

By the "ordered interchangeability" property of multiple saddle point equilibria (see Başar and Olsder, 1982), $\mu^*_{s\gamma}$ is also in equilibrium with $v^*_{s\gamma}$, but not for all $\gamma > \gamma_s$. In other words, there will exist some $\gamma \in \bar{\Gamma}_s$ for which under $\mu^*_{s\gamma}$ the upper value of the game will be infinite, even though it is finite (actually zero) if $\mu^*_{sf\gamma}$ given by (3.27) is used. We will shortly see that the pair $(\mu^*_{s\gamma}, v^*_{s\gamma})$ in fact provides a saddle-point solution to the SF slow game, but under a more restrictive condition (than (3.10)) Toward this end we now study the upper value of the slow game under the SF information pattern

To bring $L_{\gamma s}$ to standard form, we now introduce the transformations given below†

$$\begin{aligned} \hat{w}_s &= (\gamma^2 I - D_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} D_2)^{1/2} [w_s - (\gamma^2 I \\ &\quad - D_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} D_2)^{-1} \\ &\quad \cdot D_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} (B_2 u_s \\ &\quad + (A_{21} - A_{22} Q_{22}^{-1} Q_{21}) x_s)], \quad (3.29) \end{aligned}$$

$$\begin{aligned} \hat{u}_s &= \left(I + B_2'\left(A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2'\right)^{-1} B_2\right)^{1/2} \\ &\quad \cdot \left[u_s + \left(I + B_2'\left(A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2'\right)^{-1} B_2\right)^{-1}\right. \\ &\quad \cdot B_2'\left(A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2'\right)^{-1} \\ &\quad \cdot \left.(A_{21} - A_{22} Q_{22}^{-1} Q_{21}) x_s\right], \quad (3.30) \end{aligned}$$

† Note that the dependence of $\hat{w}_s$ on u_s is allowed here, because we are interested in the upper value of the game

which are valid under the condition:

$$\begin{aligned} &\gamma^2 L - D_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} D_2 > 0, \\ &\Leftrightarrow A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2' > 0. \quad (3.31) \end{aligned}$$

These transformations essentially "complete the squares" in $L_{\gamma s}$ so as to cancel out the cross terms in x_s, u_s and w_s. The transformed game then is

$$\dot{x}_s = \hat{A}_0 x_s + \hat{B}_0 \hat{u}_s + \hat{D}_0 \hat{w}_s; \quad x_s(0) = 0, \quad (3.32)$$

$$L_{\gamma s} = |x_s(t_f)|^2_{Q_{1f}} + \int_0^{t_f} (|x_s|^2_{\hat{Q}} + |\hat{u}_s|^2 - |\hat{w}_s|^2)\, dt, \quad (3.33)$$

where the coefficient matrices are (see Appendix A for details of the derivations)

$$\begin{aligned} \hat{A}_0(\gamma) &= A_{11} - A_{12} Q_{22}^{-1} Q_{21} - (S_{12} + A_{12} Q_{22}^{-1} A_{22}') \\ &\quad \cdot (S_{22} + A_{22} Q_{22}^{-1} A_{22}')^{-1} \\ &\quad (A_{21} - A_{22} Q_{22}^{-1} Q_{21}), \quad (3.34) \end{aligned}$$

$$\begin{aligned} &\hat{B}_0(\gamma) \\ &= \left(B_0 + \frac{1}{\gamma^2} D_0 D_2'\left(A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2'\right)^{-1} B_2\right) \\ &\quad \cdot \left(I + B_2'\left(A_{22} Q_{22}^{-1} A_{22}' - \frac{1}{\gamma^2} D_2 D_2'\right)^{-1} B_2\right)^{-1/2} \quad (3.35) \end{aligned}$$

$$\hat{D}_0(\gamma) = D_0(\gamma^2 I - D_2' A_{22}'^{-1} Q_{22} A_{22}^{-1} D_2)^{-1/2}$$

$$\begin{aligned} \hat{Q}(\gamma) &= Q_{11} - Q_{12} Q_{22}^{-1} Q_{21} + (A_{21}' - Q_{12} Q_{22}^{-1} A_{22}') \\ &\quad \cdot (S_{22} + A_{22} Q_{22}^{-1} A_{22}')^{-1} \\ &\quad \cdot (A_{21} - A_{22} Q_{22}^{-1} Q_{21}) \quad (3.37) \end{aligned}$$

The GRDE associated with this differential game is

$$\dot{\hat{Z}}_s + \hat{A}_0' \hat{Z}_s + \hat{Z}_s \hat{A}_0 - \hat{Z}_s \hat{S}_0 \hat{Z}_s + \hat{Q} = 0; \quad \hat{Z}_s(t_f) = Q_{1f}, \quad (3.38)$$

where

$$\hat{S}_0 := \hat{B}_0 \hat{B}_0' - \hat{D}_0 \hat{D}_0' \quad (3.39)$$

Some matrix manipulations, details of which are again included in Appendix A, show that

$$\hat{A}_0 \equiv \bar{A}_0 \quad \hat{Q} \equiv \bar{Q} \quad \hat{S}_0 \equiv S_0,$$

where $\bar{A}_0$, $\bar{Q}$, S_0 were defined earlier by (3.15), (3.18) and (3.21), respectively, and hence that the GRDE (3.38) is identical with the GRDE (3.19). It can further be shown that condition (3.31) implies condition (3.11), and hence that the set

$$\hat{\Gamma}_s := \{\gamma' > 0 : \forall \gamma \ge \gamma',\ (3.31) \text{ holds and } (3.38) \text{ has a bounded nonnegative definite solution over } [0, t_f]\}, \quad (3.40)$$

is included in $\hat{\Gamma}_s$. Now, define

$$\hat{\gamma}_s := \inf\{\gamma \in \hat{\Gamma}_s\}, \qquad (3.41)$$

which is clearly not smaller than γ_s, defined by (3.23). Following the earlier reasoning, the slow differential game with the SF information pattern has a bounded upper value if $\gamma > \hat{\gamma}_s$, and only if $\gamma \geqq \hat{\gamma}_s$. For $\gamma > \hat{\gamma}_s$, the game admits a strongly time consistent (feedback) saddle-point solution, given by the policies (for the transformed game):

$$\hat{u}^*_{s\gamma} = \hat{\mu}^*_{s\gamma}(t, x_s(t)) = -\hat{B}_0' Z_{s\gamma} x_s(t), \qquad (3.42)$$

$$\hat{w}^*_{s\gamma} = \hat{\nu}^*_{s\gamma}(t, x_s(t)) = \hat{D}_0' Z_{s\gamma} x_s(t), \qquad (3.43)$$

where $Z_{s\gamma}$ is the unique nonnegative definite solution of (3.19) (or equivalently (3.38)). Now applying the inverse transformation of (3.30) to (3.42), we obtain precisely the policy (3.28). This shows that for the slow game with SF controllers, (3.28) leads to a finite upper value, provided that $\gamma > \hat{\gamma}_s$. In general, $\hat{\gamma}_s > \gamma_s$, and hence one natural question is whether one could have a correction term, depending only on the state, added on to (3.28), which would improve the performance from $\hat{\gamma}_s$ to γ_s. The answer to this lies in the analysis of the "fast game" which is studied next, and is provided fully in Section 4.

3.2. *The fast subsystem and the associated soft-constrained game*

Let $x_f := x_2 - \bar{x}_2$, $w_f := u - u_s$, $w_f := w - w_s$ and $\tau = (t' - t/\epsilon)$ where we take t to be frozen, and t' to vary on the same scale as t. We define the fast subsystem and the associated cost (as in the standard regulator problem; see Chow and Kokotović, 1976) by:

$$\frac{\mathrm{d}}{\mathrm{d}\tau} x_f' = A_{22}(t) x_f' + B_2(t) u_f' + D_2(t) w_f';$$

$$x_f'(0) = x_f(t), \qquad (3.44)$$

$$L_{\gamma f}' = \int_0^{\infty} (|x_f'|^2_{Q_{22}(t)} + |u_f'|^2 - \gamma^2 |w_f'|^2)\, \mathrm{d}\tau. \qquad (3.45)$$

The GARE associated with this infinite-horizon game, for each t, is:

$$A_{22}'(t) Z_f + Z_f A_{22}(t) + Q_{22}(t) - Z_f S_{22}(t) Z_f = 0. \qquad (3.46)$$

We now let γ_f^t denote the minimax disturbance attenuation bound for the H^∞-optimal control problem defined by (3.44)–(3.45) under closed-loop information†, and γ_{0f}^t denote the same under open-loop information, where we must have $\gamma_{0f}^t \geq \gamma_f^t$ (see Başar and Bernhard, 1991). For every $\gamma > \gamma_f^t$, let $Z_{f\gamma}(t)$ be the minimal positive definite solution for (3.46). Then, for any $\gamma > \gamma_f^t$, the feedback controller that attains the finite upper value of the fast game is

$$u_{f\gamma}^{t*}(\tau) = \mu_{f\gamma}^{t*}(x_f^t(\tau)) = -B_2'(t) Z_{f\gamma}(t) x_f^t(\tau),$$

$$\Rightarrow \mu_{f\gamma}^*(t) = \mu_{f\gamma}^{t*}(x_f^t(0)) = -B_2'(t) Z_{f\gamma}(t) x_f(t)$$

$$= -B_2'(t) Z_{f\gamma}(t)(x_2(t) - \bar{x}_2(t)). \qquad (3.47)$$

Substitute (3.1), (3.28) and (3.26) into (3.47), to obtain

$$\begin{aligned}\mu_{f\gamma}^*(t, x(t)) = {} & -B_2' Z_{f\gamma} x_2 - B_2' Z_{f\gamma} Q_{22}^{-1}(A_{12}' Z_{s\gamma} \\ & + Q_{21} - A_{22}'(S_{22} + A_{22} Q_{22}^{-1} A_{22}')^{-1} \\ & \cdot ((S_{21} + A_{22} Q_{22}^{-1} A_{12}') Z_{s\gamma} \\ & - (A_{21} - A_{22} Q_{22}^{-1} Q_{21}))) x_1(t). \qquad (3.48)\end{aligned}$$

We now define

$$\gamma_f := \sup_{t \in [0, t_f]} \gamma_f^t, \qquad (3.49)$$

$$\gamma_{0f} := \sup_{t \in [0, t_f]} \gamma_{0f}^t. \qquad (3.50)$$

Then, for every $\gamma > \gamma_f$, the GARE (3.46) admits a positive definite solution for all $t \in [0, t_f]$. Let

$$\bar{\gamma} := \max\{\gamma_f, \gamma_s\}. \qquad (3.51)$$

We will shortly see that this value plays an important role in our problem.

3.3. *The infinite-horizon case*

We now turn to the infinite horizon case. Let A, B, D and Q be time invariant and Q_f be zero. By following steps similar to those in the finite horizon case, we first decompose the system into slow and fast subsystems. After an appropriate transformation, the slow game with FI controller is described by

$$\dot{x}_s = \bar{A}_0 x_s + \bar{B}_0 \bar{u}_s + \bar{D}_0 \bar{w}_s; \quad x_s(0) = 0, \qquad (3.52)$$

$$L_{\gamma s} = \int_0^{\infty} (|x_s|^2_{\bar{Q}} + |\bar{u}_s|^2 - |\bar{w}_s|^2)\, \mathrm{d}t, \qquad (3.53)$$

where $\bar{A}_0$, $\bar{B}_0$, $\bar{D}_0$ and $\bar{Q}$ are as defined before, with the only difference being that they are now time invariant. The associated GARE is

$$\bar{A}_0' \bar{Z}_s + \bar{Z}_s \bar{A}_0 - \bar{Z}_s S_0 \bar{Z}_s + \bar{Q} = 0. \qquad (3.54)$$

Let us define the following set as the counterpart of (3.22):

$$\begin{aligned}\bar{\Gamma}_{s\infty} := \{\gamma' > 0 : \forall \gamma \geq \gamma',\ & S_{22} + A_{22} Q_{22}^{-1} A_{22}' > 0, \\ & \text{(3.54) has a nonnegative definite solution} \\ & Z_{s\gamma}, \text{ and } \bar{A}_0'(\gamma) - S_0(\gamma) Z_{s\gamma} \text{ is Hurwitz}\}\end{aligned}$$

and further define

$$\gamma_{s\infty} := \inf\{\gamma \in \bar{\Gamma}_{s\infty}\}. \qquad (3.55)$$

† To ensure that $\gamma_f^t < \infty$, it will be sufficient to take the pair $(A_{22}(t), B_2(t))$ to be stabilizable.

For every $\gamma > \gamma_{s\infty}$†, let $Z_{s\gamma}$ be the minimal positive definite solution to (3.54). Then, for each $\gamma > \gamma_{s\infty}$, the FI controller that attains the upper value is the time-invariant version of (3.24), and the maximizing disturbance with respect to this control is the time-invariant version of (3.25). Then we can solve for $\mu^*_{s\gamma}$ and $v^*_{s\gamma}$, which are the time-invariant versions of (3.26) and (3.27), respectively‡. To obtain the corresponding SF controller, we substitute $v^*_{s\gamma}$ into $\mu^*_{s1\gamma}$, to arrive at an expression for $\mu^*_{s\gamma}$, which is the time-invariant version of (3.28). Note that this SF controller yields a finite upper value if

$$\gamma > \hat{\gamma}_{s\infty} := \inf\{\gamma \in \hat{\Gamma}_{s\infty}\},$$

where

$$\hat{\Gamma}_{s\infty} := \{\gamma' \geq \gamma_{s\infty} : \forall \gamma > \gamma',\ (3.31) \text{ holds}\}.$$

The fast part of the system is the same as in the finite-horizon case, where the coefficient matrices are now constants. The fast game is described by (3.44)–(3.45), and the GARE is the same as (3.46). We will use $\gamma_{f\infty}$ to denote the minimax disturbance attenuation bound (and γ_{ol}, for the open-loop case), and let $Z_{f\gamma}$ denote the minimal positive definite solution to (3.46). Then, the minimizing controller is given by (3.48). As the counterpart of (3.51), we define

$$\bar{\gamma}_\infty := \max\{\gamma_{f\infty}, \gamma_{s\infty}\}. \tag{3.56}$$

This quantity will also play an important role in our analysis in the next section.

4 MAIN RESULTS

Before we present the main results, we first give three useful results, listed below as Lemma 1, Lemma 2, and Fact 1.

Lemma 1. Consider the following GARE:

$$A'Z + ZA + Q - Z\left(BB' - \frac{1}{\gamma^2}DD'\right)Z = 0, \tag{4.1}$$

where the matrices A, Q, B and D are taken to be quite general, apart from the requirement that $Q > 0$ and (A, B) is controllable. Let $0 \leq \gamma_m < \infty$ be the infimum of all γs under which this GARE admits a nonnegative definite solution. Then, for $\gamma > \gamma_m$,

$$BB' - \frac{1}{\gamma^2}DD' + AQ^{-1}A' > 0. \tag{4.2}$$

† Again, for $\gamma_{s\infty}$ to be finite, it will be sufficient to have the pair $(\bar{A}_0, \bar{B}_0)$ controllable at $\gamma = \infty$, which is equivalent to having the pair (A_0, B_0) controllable

‡ Note that here $v^*_{s\gamma}$ is maximizing under $\mu^*_{s\gamma}$, but $\mu^*_{s\gamma}$ is not necessarily minimizing under $v^*_{s\gamma}$, see Başar and Bernhard (1991)

Proof. Since (A, Q) is observable, the nonnegative definite solution of the GARE is actually positive definite (see Başar and Bernhard, 1991). Let Z be the minimal positive definite solution of (4.1). It is well known that, under the conditions specified, the matrix $A - (BB' - (1/\gamma^2)DD')Z$ is Hurwitz. Let $S := BB' - (1/\gamma^2)DD'$. Then, using standard matrix manipulations and the GARE (4.1), we obtain:

$$\begin{aligned}
S + AQ^{-1}A' &= S - Z^{-1}(-ZA)Q^{-1}A' \\
&= S - Z^{-1}(A'Z + Q - ZSZ)Q^{-1}A' \\
&= S - Z^{-1}A' + SZQ^{-1}A' - Z^{-1}A'ZQ^{-1}A' \\
&= (S - Z^{-1}A')(I + ZQ^{-1}A') \\
&= Z^{-1}(ZS - A')ZQ^{-1}(Q + A'Z)Z^{-1} \\
&= Z^{-1}(ZS - A')ZQ^{-1}Z(SZ - A)Z^{-1} > 0.
\end{aligned}$$

Lemma 2. Under Assumptions A1 and A3, if $\gamma > \gamma_f$ (or $\gamma > \gamma_{f\infty}$, in the infinite horizon case), then $\bar{Q}$ defined by (3.18) is nonnegative definite.

Proof. Since $\gamma > \gamma_f$ ($\gamma > \gamma_{f\infty}$) and $Q_{22} > 0$, it follows from Lemma 1 that $S_{22} + A_{22}Q_{22}^{-1}A'_{22} > 0$. Since nonnegative definiteness of Q implies $Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} \geq 0$, the result follows.

Next we recall a well-known fact.

Fact 1. Let A, $P \geq 0$ and $Q \geq 0$ be three square matrices of the same dimensions. Then if (A, Q) is observable, so are $(A, Q + P)$ and $(A + KP_s, Q + P)$, the latter for all K of appropriate dimensions, where $P_s'P_s = P$.

4.1. *The composite controller*

We now introduce, for both finite and infinite horizons, the composite controller:

$$\mu^*_{c\gamma}(t, x) = \mu^*_{s\gamma}(t, x) + \mu^*_{f\gamma}(t, x), \tag{4.3}$$

where $\mu^*_{s\gamma}$ and $\mu^*_{f\gamma}$ were defined by (3.28) and (3.48), respectively, and $\gamma > \bar{\gamma}$ (or $\bar{\gamma}_\infty$, in the infinite horizon case). After some manipulations, this composite controller can be rewritten as

$$\mu^*_{c\gamma}(t, x) = -B'\begin{bmatrix} Z_{s\gamma} & 0 \\ Z_c & Z_{f\gamma} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{4.4}$$

where

$$\begin{aligned}
Z_c := {} & Z_{f\gamma}Q_{22}^{-1}(A'_{12}Z_{s\gamma} + Q_{21}) \\
& - (I + Z_{f\gamma}Q_{22}^{-1}A'_{22})(S_{22} + A_{22}Q_{22}^{-1}A'_{22})^{-1} \\
& \cdot ((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma} \\
& - (A_{21} - A_{22}Q_{22}^{-1}Q_{21})) && (4.5) \\
\equiv {} & Z_{f\gamma}U + V && (4.6) \\
= {} & Z_{f\gamma}(U_1 Z_{s\gamma} + U_2) + V_1 Z_{s\gamma} + V_2, && (4.7)
\end{aligned}$$

and

$$U_1 := Q_{22}^{-1}A_{22}' - Q_{22}^{-1}A_{22}'(S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}(S_{21} + A_{22}Q_{22}^{-1}A_{22}'), \quad (4.8)$$

$$U_2 := Q_{22}^{-1}Q_{21} + Q_{22}^{-1}A_{22}'(S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21}), \quad (4.9)$$

$$V_1 := -(S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}(S_{21} + A_{22}Q_{22}^{-1}A_{12}'), \quad (4.10)$$

$$V_2 := (S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21}). \quad (4.11)$$

Now, we are in a position to present our main results, first for the infinite horizon and subsequently for the finite horizon cases.

4.2. *The infinite horizon case*

To guarantee that $\bar{\gamma}_\infty$ is finite (although the theorem to be presented below is true also when it is infinite), we make two additional assumptions:

Assumptions

A4. $(A_0, Q_{11} - Q_{12}Q_{22}^{-1}Q_{21})$ is observable.
A5. (A_0, B_0) and (A_{22}, B_2) are controllable.

Theorem 1. Consider the singularly perturbed system (2.1)–(2.6), with $t_f = \infty$, $Q_f = 0$ and A, B, D, Q time-invariant. If Assumptions A1–A5 hold, then,

(1) $\gamma^*(\epsilon) \le \bar{\gamma}_\infty$, asymptotically as $\epsilon \to 0$, where $\bar{\gamma}_\infty$, as defined in (3.56), is finite.
(2) $\forall \gamma > \bar{\gamma}_\infty$, $\exists \epsilon_\gamma > 0$ such that $\forall \epsilon \in [0, \epsilon_\gamma)$, the GARE (2.26) admits a positive definite solution, and consequently, the game has a finite upper value. Furthermore, the minimal such solution can be approximated by

$$\bar{Z} = \begin{bmatrix} Z_{s\gamma} + O(\epsilon) & \epsilon Z_c' + O(\epsilon^2) \\ \epsilon Z_c + O(\epsilon^2) & \epsilon Z_{f\gamma} + O(\epsilon^2) \end{bmatrix}. \quad (4.12)$$

(3) $\forall \gamma > \bar{\gamma}_\infty$, if we apply the composite controller $\mu_{c\gamma}^*$ to the system, then $\exists \epsilon_\gamma' > 0$ such that, $\forall \epsilon \in [0, \epsilon_\gamma')$, the disturbance attenuation level γ is attained for the full-order system.
(4) $\forall \gamma > \max\{\gamma_{s\infty}, \gamma_{of\infty}\} = \max\{\hat{\gamma}_{s\infty}, \gamma_{of\infty}\}$, if we only apply the slow controller $\mu_{s\gamma}^*$ to the system, then $\exists \epsilon_\gamma'' > 0$ such that, $\forall \epsilon \in [0, \epsilon_\gamma'')$, the disturbance attenuation level γ is attained for the full-order system.

Proof. We first note that under Assumptions A3, A4 and A5, both slow (with either SF or FI controllers) and fast games have saddle-point solutions for sufficiently large values of γ, since at $\gamma = \infty$ both problems become "regular" LQR problems admitting stabilizing optimal controllers. This shows that both $\gamma_{s\infty}$ and $\gamma_{f\infty}$ are finite, implying that $\bar{\gamma}_\infty$ is also finite. Now fix $\gamma > \bar{\gamma}_\infty$. Under Assumptions A1–A4, we know from Lemma 1 that $S_{22} + A_{22}Q_{22}^{-1}A_{22}' > 0$. Then, by Lemma 2, $\bar{Q} \ge 0$, and also by the nonnegative definiteness of Q, $Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} \ge 0$. Since

$$Q_s := [I \quad -A_{21}'A_{22}'^{-1}] \begin{bmatrix} Q_{11}(t) & Q_{12}(t) \\ Q_{21}(t) & Q_{22}(t) \end{bmatrix} \times \begin{bmatrix} I \\ -A_{22}^{-1}A_{21} \end{bmatrix}$$
$$= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (Q_{12}Q_{22}^{-1} - A_{21}'A_{22}'^{-1})Q_{22}(Q_{22}^{-1}Q_{21} - A_{22}^{-1}A_{21}),$$

by Fact 1, (A_0, Q_s) is observable. Also, we can rewrite (3.15) as follows:

$$\bar{A}_0 = A_0 - [B_0 \quad D_0] \begin{bmatrix} B_2' \\ -(1/\gamma^2)D_2' \end{bmatrix} \times (S_{22} + A_{22}Q_{22}^{-1}A_{22}')^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21}).$$

By Fact 1, $(\bar{A}_0, \bar{Q})$ is observable. From the observability of pairs $(\bar{A}_0, \bar{Q})$ and (A_{22}, Q_{22}), we conclude that $Z_{s\gamma} > 0$, $Z_{f\gamma} > 0$ (see Başar and Bernhard, 1991). Furthermore, $A_{22} - S_{22}Z_{f\gamma}$ is Hurwitz for $\gamma > \gamma_{f\infty}$, and by definition $\bar{A}_0 - S_0 Z_{s\gamma}$ is Hurwitz, for $\gamma > \gamma_{s\infty}$. From the observability of (A_{22}, Q_{22}), we can deduce that, for ϵ sufficiently small, the pair (A_ϵ, Q) is also observable (see Kokotović *et al.*, 1986). Then the game (2.1)–(2.6) has a bounded upper value if (2.26) admits a positive definite solution, while the upper value is unbounded if there is no nonnegative definite solution. We show in Appendix C that the set of coupled matrix equations (2.30)–(2.32) admits a solution $Z_{11} = Z_{s\gamma}$, $Z_{12} = Z_c'$ and $Z_{22} = Z_{f\gamma}$. We now use the Implicit Function Theorem to complete the proofs of parts 1 and 2. Let $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$ be the vector forms of $\bar{Z}_{11}$, $\bar{Z}_{12}$ and $\bar{Z}_{22}$, respectively. Let us write equation (2.27)–(2.29) as $\xi(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = 0$. We already know that $\xi(0, K_1, K_2, K_3) = 0$, where K_1, K_2 and K_3 are the vector forms of $Z_{s\gamma}$, Z_c and $Z_{f\gamma}$. Let

$$\bar{\Xi}_{ij} = \left.\frac{\partial \xi_i}{\partial \bar{K}_j}\right|_{(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = (0, K_1, K_2, K_3)} \qquad i = 1, 2, 3, j = 1, 2, 3.$$

Then

$$\bar{\Xi}_{31} = 0; \quad \bar{\Xi}_{32} = 0,$$

$$\bar{\Xi}_{33} = I_{n_2} \otimes (A_{22} - S_{22}Z_{f\gamma}) + (A_{22}' - Z_{f\gamma}S_{22}) \otimes I_{n_2},$$

$$\bar{\Xi}_{22} = (A_{22}' - Z_{f\gamma}S_{22}) \otimes I_{n1},$$

where $\otimes$ refers to the Kronecker product. Furthermore, $\bar{\Xi}_{33}$ and $\bar{\Xi}_{22}$ are invertible. Solving

for $\bar{Z}_{12}$ from (2.28) we obtain

$$\bar{Z}_{12} = -\bar{Z}_{11}(A_{12} - S_{12}\bar{Z}_{22})(A_{22} - S_{22}\bar{Z}_{22})^{-1} - (Q_{12} + A'_{21}\bar{Z}_{22})(A_{22} - S_{22}\bar{Z}_{22})^{-1} + O(\epsilon), \quad (4.13)$$

and its substitution into (2.27) leads to an equation in terms of ϵ, $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$, which we denote as $\xi_1(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = 0$. Let

$$\Xi_{1j} = \frac{\partial \xi_1}{\partial \bar{K}_j}\bigg|_{(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = (0, \bar{K}_1, \bar{K}_2, \bar{K}_3)}, \quad j = 1, 2, 3,$$

then

$$\Xi_{12} = 0, \qquad \Xi_{11} = I_{n_1} \otimes R + R' \otimes I_{n_1}, \quad (4.14)$$

where $R = \bar{A}_0 - S_0 Z_{s\gamma}$ (see Appendix C). Hence,

$$\begin{bmatrix} \Xi_{11} & \Xi_{12} & \Xi_{13} \\ \bar{\Xi}_{21} & \bar{\Xi}_{22} & \bar{\Xi}_{23} \\ \bar{\Xi}_{31} & \bar{\Xi}_{32} & \bar{\Xi}_{33} \end{bmatrix} = \begin{bmatrix} \Xi_{11} & 0 & * \\ * & \bar{\Xi}_{22} & * \\ 0 & 0 & \Xi_{33} \end{bmatrix} \text{ is nonsingular}$$

where $*$ refers to any constant matrix. This implies that

$$\begin{bmatrix} \bar{\Xi}_{11} & \bar{\Xi}_{12} & \bar{\Xi}_{13} \\ \bar{\Xi}_{21} & \bar{\Xi}_{21} & \bar{\Xi}_{23} \\ \bar{\Xi}_{31} & \bar{\Xi}_{32} & \bar{\Xi}_{33} \end{bmatrix} \text{ is also nonsingular}$$

Then, by the Implicit Function Theorem, $\exists \epsilon_1 > 0$ such that $\forall \epsilon \in [0, \epsilon_1)$, there exist $\bar{Z}_{11}(\epsilon)$, $\bar{Z}_{12}(\epsilon)$ and $\bar{Z}_{22}(\epsilon)$ that solve (2.27)–(2.29) and $\bar{Z}_{11}(\epsilon) = Z_{s\gamma} + O(\epsilon)$, $\bar{Z}_{12}(\epsilon) = Z'_c + O(\epsilon)$ and $\bar{Z}_{22}(\epsilon) = Z_{f\gamma} + O(\epsilon)$. Since $Z_{s\gamma} > 0$, $Z_{f\gamma} > 0$ and $Z_{s\gamma}$, $Z_{f\gamma}$ are the minimal positive definite solutions for (3.54), (3.46), $\exists \epsilon_\gamma$, $0 < \epsilon_\gamma < \epsilon_1$, such that $\forall \epsilon \in [0, \epsilon_\gamma]$, the minimal positive definite solution† for (2.27)–(2.29) is

$$\bar{Z}(\epsilon) = \begin{bmatrix} \bar{Z}_{11}(\epsilon) & \epsilon\bar{Z}_{12}(\epsilon) \\ \epsilon\bar{Z}_{21}(\epsilon) & \epsilon\bar{Z}_{22}(\epsilon) \end{bmatrix} > 0$$

So far, we have proved parts 1 and 2. For part 3, we use $\mu^*_{c\gamma}$ in system (2.1)–(2.6)

$$\begin{bmatrix} \dot{x}_1 \\ \epsilon\dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} - B_1B'_1Z_{s\gamma} - B_1B'_2Z_c & A_{12} - B_1B'_2Z_{f\gamma} \\ A_{21} - B_2B'_1Z_{s\gamma} - B_2B'_2Z_c & A_{22} - B_2B'_2Z_{f\gamma} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} w, \quad (4.15)$$

$$L_\gamma = \int_0^\infty (|x|^2_Q - \gamma^2 |w|^2)\, dt \quad (4.16)$$

The ARE associated with this maximization problem is

$$\tilde{A}'_\epsilon \tilde{W} + \tilde{W}\tilde{A}_\epsilon + \frac{1}{\gamma^2}\tilde{W}D_\epsilon D'_\epsilon\tilde{W} + \bar{Q} = 0, \quad (4.17)$$

where

$$\tilde{A}_\epsilon = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ (1/\epsilon)\tilde{A}_{21} & (1/\epsilon)\tilde{A}_{22} \end{bmatrix}, \quad (4.18)$$

$$\tilde{A}_{11} = A_{11} - B_1B'_1Z_{s\gamma} - B_1B'_2Z_c, \quad (4.19)$$

$$\tilde{A}_{12} = A_{12} - B_1B'_2Z_{f\gamma}, \quad (4.20)$$

$$\tilde{A}_{21} = A_{21} - B_2B'_1Z_{s\gamma} - B_2B'_2Z_c, \quad (4.21)$$

$$\tilde{A}_{22} = A_{22} - B_2B'_2Z_{f\gamma}, \quad (4.22)$$

$$\bar{Q} = Q + \begin{bmatrix} Z_{s\gamma} & Z'_c \\ 0 & Z_{f\gamma} \end{bmatrix}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}[B'_1 \quad B'_2] \times \begin{bmatrix} Z_{s\gamma} & 0 \\ Z_c & Z_{f\gamma} \end{bmatrix} \quad (4.23)$$

$$= \begin{bmatrix} \bar{Q}_{11} & \bar{Q}_{12} \\ \bar{Q}_{21} & \bar{Q}_{22} \end{bmatrix} \quad (4.24)$$

By Fact 1, the observability of (A_ϵ, Q) implies the observability of $(\tilde{A}_\epsilon, \bar{Q})$. Then the maximization problem (4.15)–(4.16) has a finite cost if (4.17) admits a positive definite solution, and the cost is unbounded if (4.17) does not have a nonnegative definite solution. Suppose that $\tilde{W}$ is in the form

$$\tilde{W} = \begin{bmatrix} \tilde{W}_{11} & \epsilon\tilde{W}_{12} \\ \epsilon\tilde{W}_{21} & \epsilon\tilde{W}_{22} \end{bmatrix},$$

where $\tilde{W}_{12} = \tilde{W}'_{21}$. Expanding (4.17) around $\epsilon = 0$ we obtain

$$\tilde{A}'_{11}\tilde{W}_{11} + \tilde{A}'_{21}\tilde{W}'_{12} + \tilde{W}_{11}\tilde{A}_{11} + \tilde{W}_{12}\tilde{A}_{21} + \bar{Q}_{11} + \frac{1}{\gamma^2}(\tilde{W}_{11}D_1D'_1\tilde{W}_{11} + \tilde{W}_{12}D_2D'_2\tilde{W}_{11} + \tilde{W}_{11}D_1D'_2\tilde{W}'_{12} + \tilde{W}_{12}D_2D'_2\tilde{W}'_{12}) = 0, \quad (4.25)$$

† The positive definiteness follows from $Z_{s\gamma} > 0$, $Z_{f\gamma} > 0$ and the following well known result: let

$$Z = \begin{bmatrix} Q & P \\ P' & R \end{bmatrix}$$

where Q and R are positive definite matrices. Then $Z > 0$ if and only if $Q - PR^{-1}P' > 0$.

$$\epsilon\bar{A}'_{11}\bar{W}_{12}+\bar{A}'_{21}\bar{W}_{22}+\bar{W}_{11}\bar{A}_{12}+\bar{W}_{12}\bar{A}_{22}+\bar{Q}_{12}+\frac{1}{\gamma^2}(\epsilon\bar{W}_{11}D_1D_1'\bar{W}_{12}+\epsilon\bar{W}_{12}D_2D_1'\bar{W}_{12}+\bar{W}_{11}D_1D_2'\bar{W}_{22}+\bar{W}_{12}D_2D_2'\bar{W}_{22})=0, \quad (4.26)$$

$$\epsilon\bar{A}'_{12}\bar{W}_{12}+\bar{A}'_{22}\bar{W}_{22}+\epsilon\bar{W}'_{12}\bar{A}_{12}+\bar{W}_{22}\bar{A}_{22}+\bar{Q}_{22}+\frac{1}{\gamma^2}(\epsilon^2\bar{W}'_{12}D_1D_1'\bar{W}_{12}+\epsilon\bar{W}_{22}D_2D_1'\bar{W}_{12}+\epsilon\bar{W}'_{12}D_1D_2'\bar{W}_{22}+\bar{W}_{22}D_2D_2'\bar{W}_{22})=0. \quad (4.27)$$

It is straightforward to verify that when $\epsilon=0$, $\bar{W}_{11}=Z_{s\gamma}$, $\bar{W}_{12}=Z'_c$ and $\bar{W}_{22}=Z_{f\gamma}$ satisfy (4.25)–(4.27). Let $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$ be the vector forms of $\bar{W}_{11}$, $\bar{W}_{12}$ and $\bar{W}_{22}$, respectively. Viewing equation (4.25)–(4.27) as the vector equation $\bar{\eta}(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=0$, we let K_1, K_2 and K_3 be the vector forms of $Z_{s\gamma}$, Z_c and $Z_{c\gamma}$, and introduce

$$\bar{\Xi}_{ij}=\frac{\partial\bar{\eta}_i}{\partial\bar{K}_j}\bigg|_{(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=(0,K_1,K_2,K_3)} \qquad i=1,2,3 \quad j=1,2,3.$$

Then

$$\bar{\Xi}_{31}=0,\quad \bar{\Xi}_{32}=0,$$

$$\bar{\Xi}_{33}=I_{n_2}\otimes(A_{22}-S_{22}Z_{f\gamma})+(A'_{22}-Z_{f\gamma}S_{22})\otimes I_{n_2},$$

$$\bar{\Xi}_{22}=(A'_{22}-Z_{f\gamma}S_{22})\otimes I_{n_1},$$

which shows that $\bar{\Xi}_{33}$ and $\bar{\Xi}_{22}$ are nonsingular. Solving for $\bar{W}_{12}$ from equation (4.26) we obtain:

$$\bar{W}_{12}=-\bar{W}_{11}\Big(\bar{A}_{12}+\frac{1}{\gamma^2}D_1D_2'\bar{W}_{22}\Big)\times\Big(\bar{A}_{22}+\frac{1}{\gamma^2}D_2D_2'\bar{W}_{22}\Big)^{-1}-(\bar{Q}_{12}+\bar{A}'_{21}\bar{W}_{22})\Big(\bar{A}_{22}+\frac{1}{\gamma^2}D_2D_2'\bar{W}_{22}\Big)^{-1}+O(\epsilon). \quad (4.28)$$

Substitution of this into (4.25) leads to an equation in terms of ϵ, $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$, which we write as $\eta_1(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=0$. Let

$$\Xi_{1j}=\frac{\partial\eta_1}{\partial\bar{K}_j}\bigg|_{(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=(0,K_1,K_2,K_3)} \qquad j=1,2,3.$$

Then, we have the structure:

$$\Xi_{12}=0,\qquad \Xi_{11}=I_{n_1}\otimes R+R'\otimes I_{n_1}, \quad (4.29)$$

where we can readily see, from the analysis of Appendix C, that $R=\bar{A}_0-S_0Z_{s\gamma}$. Then Ξ_{11} is invertible by the stability of R, and thus 3 is established by the Implicit Function Theorem.

For part 4, we first compare (4.10) and (4.11) with (3.28), to obtain $\mu^*_{s\gamma}=-(B_1'Z_{s\gamma}+B_2'V)x_s(t)$. Substituting $\mu^*_{s\gamma}$ into system (2.1)–(2.6), we arrive at the maximization problem:

$$\begin{bmatrix}\dot{x}_1\\ \epsilon\dot{x}_2\end{bmatrix}=\begin{bmatrix}A_{11}-B_1B_1'Z_{s\gamma}-B_1B_2'V & A_{12}\\ A_{21}-B_2B_1'Z_{s\gamma}-B_2B_2'V & A_{22}\end{bmatrix}\times\begin{bmatrix}x_1\\ x_2\end{bmatrix}+\begin{bmatrix}D_1\\ D_2\end{bmatrix}w, \quad (4.30)$$

$$J=\int_0^\infty(|x|_Q^2-\gamma^2|w|^2)\,dt\to\text{maximize}. \quad (4.31)$$

The associated ARE is

$$\hat{A}_\epsilon'\bar{Y}+\bar{Y}\hat{A}_\epsilon+\frac{1}{\gamma^2}\bar{Y}D_\epsilon D_\epsilon'\bar{Y}+\hat{Q}=0, \quad (4.32)$$

where

$$\hat{A}_\epsilon=\begin{bmatrix}\hat{A}_{11} & \hat{A}_{12}\\ (1/\epsilon)\hat{A}_{21} & (1/\epsilon)\hat{A}_{22}\end{bmatrix}, \quad (4.33)$$

$$\hat{A}_{11}=A_{11}-B_1B_1'Z_{s\gamma}-B_1B_2'V, \quad (4.34)$$

$$\hat{A}_{12}=A_{12}, \quad (4.35)$$

$$\hat{A}_{21}=A_{21}-B_2B_1'Z_{s\gamma}-B_2B_2'V, \quad (4.36)$$

$$\hat{A}_{22}=A_{22}, \quad (4.37)$$

$$\hat{Q}=Q+\begin{bmatrix}Z_{s\gamma} & V'\\ 0 & 0\end{bmatrix}\begin{bmatrix}B_1\\ B_2\end{bmatrix}\begin{bmatrix}Z_{s\gamma} & 0\\ V & 0\end{bmatrix} \quad (4.38)$$

$$=\begin{bmatrix}\hat{Q}_{11} & \hat{Q}_{12}\\ \hat{Q}_{21} & \hat{Q}_{22}\end{bmatrix}. \quad (4.39)$$

By Fact 1, the observability of (A_ϵ, Q) implies the observability of $(\hat{A}_\epsilon, \hat{Q})$. Then the optimization problem (4.30)–(4.31) has a finite cost if (4.32) admits a positive definite solution and the cost is not bounded if (4.32) does not have a nonnegative definite solution. Also, $\gamma>\gamma_{of*}$ implies that $-(1/\gamma)^2D_2D_2'+A_{22}Q_{22}^{-1}A_{22}'>0$ by Lemma 1, and consequently $S_{22}+A_{22}Q_{22}^{-1}A_{22}'>0$. Thus, the equality $\max\{\gamma_{s*},\gamma_{of*}\}=\max\{\hat{\gamma}_{s*},\gamma_{of*}\}$ holds by the definitions of γ_{s*} and $\hat{\gamma}_{s*}$. Starting again with the form:

$$\bar{Y}=\begin{bmatrix}\bar{Y}_{11} & \epsilon\bar{Y}_{12}\\ \epsilon\bar{Y}_{21} & \epsilon\bar{Y}_{22}\end{bmatrix},$$

where $\bar{Y}_{12}=\bar{Y}'_{21}$, we expand (4.32) around $\epsilon=0$:

$$\hat{A}'_{11}\bar{Y}_{11}+\hat{A}'_{21}\bar{Y}'_{12}+\bar{Y}_{11}\hat{A}_{11}+\bar{Y}_{12}\hat{A}_{21}+\hat{Q}_{11}+\frac{1}{\gamma^2}(\bar{Y}_{11}D_1D_1'\bar{Y}_{11}+\bar{Y}_{12}D_2D_1'\bar{Y}_{11}+\bar{Y}_{11}D_1D_2'\bar{Y}'_{12}+\bar{Y}_{12}D_2D_2'\bar{Y}'_{12})=0 \quad (4.40)$$

$$\epsilon\hat{A}'_{11}\bar{Y}_{12}+\hat{A}'_{21}\bar{Y}_{22}+\bar{Y}_{11}\hat{A}_{12}+\bar{Y}_{12}\hat{A}_{22}+\hat{Q}_{12}+\frac{1}{\gamma^2}(\epsilon\bar{Y}_{11}D_1D_1'\bar{Y}_{12}+\epsilon\bar{Y}_{12}D_2D_1'\bar{Y}_{11}+\bar{Y}_{11}D_1D_2'\bar{Y}_{22}+\bar{Y}_{12}D_2D_2'\bar{Y}_{22})=0, \quad (4.41)$$

$$\epsilon\hat{A}'_{12}\bar{Y}_{12} + \hat{A}'_{22}\bar{Y}_{22} + \epsilon\bar{Y}'_{12}\hat{A}_{12} + \bar{Y}_{22}\hat{A}_{22} + \hat{Q}_{22} + \frac{1}{\gamma^2}(\epsilon^2\bar{Y}'_{12}D_1D'_1\bar{Y}_{12} + \epsilon\bar{Y}_{22}D_2D'_1\bar{Y}_{12} + \epsilon\bar{Y}'_{12}D_1D'_2\bar{Y}_{22} + \bar{Y}_{22}D_2D'_2\bar{Y}_{22}) = 0. \quad (4.42)$$

It is shown in Appendix C that for $\epsilon = 0$, $\bar{Y}_{11} = Z_{s\gamma}$, $\bar{Y}_{21} = Z_{of\gamma}U + V$ and $\bar{Y}_{22} = Z_{of\gamma}$ satisfy (4.40)–(4.42), where $Z_{of\gamma}$ denotes the solution to the "open-loop" GARE of the fast subgame:

$$A'_{22}Z_f + Z_fA_{22} + Z_f\frac{1}{\gamma^2}D_2D'_2Z_f + Q_{22} = 0.$$

We now show how the Implicit Function Theorem can again be employed to prove 4. Let $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$ be the vector forms of $\bar{Y}_{11}$, $\bar{Y}_{12}$ and $\bar{Y}_{22}$, respectively. Writing equations (4.40)–(4.42) as $\bar{\lambda}(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = 0$, we let K_1, K_2 and K_3 to be the vector forms of $Z_{s\gamma}$, $V' + U'Z_{of\gamma}$ and $Z_{of\gamma}$. Introducing

$$\bar{\Xi}_{ij} = \left.\frac{\partial\bar{\lambda}_i}{\partial\bar{K}_j}\right|_{(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=(0,K_1,K_2,K_3)}, \quad i = 1, 2, 3,\ j = 1, 2, 3,$$

we arrive at

$$\bar{\Xi}_{31} = 0; \quad \bar{\Xi}_{32} = 0,$$

$$\bar{\Xi}_{33} = I_{n_2} \otimes \left(A_{22} + \frac{1}{\gamma^2}D_2D'_2Z_{of\gamma}\right) + \left(A'_{22} + \frac{1}{\gamma^2}Z_{of\gamma}D_2D'_2\right) \otimes I_{n_2},$$

$$\bar{\Xi}_{22} = \left(A'_{22} + \frac{1}{\gamma^2}Z_{of\gamma}D_2D'_2\right) \otimes I_{n_1}.$$

Since $\gamma > \gamma_{o f\gamma}$, $A_{22} + (1/\gamma^2)D_2D'_2Z_{of\gamma}$ is Hurwitz, and hence $\bar{\Xi}_{33}$ and $\bar{\Xi}_{22}$ are invertible. From equation (4.41):

$$\bar{Y}_{12} = -\bar{Y}_{11}\left(A_{12} + \frac{1}{\gamma^2}D_1D'_1\bar{Y}_{22}\right) \times \left(A_{22} + \frac{1}{\gamma^2}D_2D'_2\bar{Y}_{22}\right)^{-1} - (Q_{12} + \hat{A}'_{21}\bar{Y}_{22}) \times \left(A_{22} + \frac{1}{\gamma^2}D_2D'_2\bar{Y}_{22}\right)^{-1} + O(\epsilon). \quad (4.43)$$

Substitution of (4.43) into (4.40) leads to an equation in terms of ϵ, $\bar{K}_1$, $\bar{K}_2$ and $\bar{K}_3$, which we write as $\lambda_1(\epsilon, \bar{K}_1, \bar{K}_2, \bar{K}_3) = 0$. Let

$$\bar{\Xi}_{1j} = \left.\frac{\partial\lambda_1}{\partial\bar{K}_j}\right|_{(\epsilon,\bar{K}_1,\bar{K}_2,\bar{K}_3)=(0,K_1,K_2,K_3)}, \quad j = 1, 2, 3.$$

Then, we have the structure

$$\Xi_{12} = 0, \qquad \Xi_{11} = I_{n_1} \otimes R + R' \otimes I_{n_1}, \quad (4.44)$$

where $R = \bar{A}_0 - S_0Z_{s\gamma}$. (See Appendix C for the manipulations that lead to this expression for R.)

Hence,

$$\begin{bmatrix} \Xi_{11} & \Xi_{12} & \Xi_{13} \\ \bar{\Xi}_{21} & \bar{\Xi}_{22} & \bar{\Xi}_{23} \\ \bar{\Xi}_{31} & \bar{\Xi}_{32} & \bar{\Xi}_{33} \end{bmatrix} = \begin{bmatrix} \Xi_{11} & 0 & * \\ * & \bar{\Xi}_{22} & * \\ 0 & 0 & \bar{\Xi}_{33} \end{bmatrix} \text{ is nonsingular,}$$

where $*$ refers to any constant matrix. Part 4 of the theorem then follows from the implicit Function Theorem. □

Remark 1. Although the theorem above provides only an upper bound for the H^∞-optimal performance level of the full-order plant, we have actually proved recently, in Pan and Başar (1992), that this upper bound is also a lower bound for the performance level for sufficiently small ϵ. Thus, we have the following limit.†

$$\lim_{\epsilon\to 0+} \gamma^*_\epsilon(\epsilon) = \bar{\gamma}_*. \quad (4.45)$$

4.3 *The finite horizon case*

Theorem 2. For the singularly perturbed system (2.1)–(2.6), let Assumptions A1–A3 be satisfied, the pair $(A_{22}(t), B_2(t))$ be controllable for each $t \in [0, t_f]$, and the following condition hold:

$Q_{f22} \le Z_{f\gamma}(t_f)$‡, where $Z_{f\gamma}(t_f)$ is the solution to (3.46) at $t = t_f$ with γ fixed.

Then,

(1) $\gamma^*(\epsilon) \le \bar{\gamma}$, asymptotically as $\epsilon \to 0$, where $\bar{\gamma}$, as defined in (3.51), is finite.

(2) $\forall\gamma > \bar{\gamma}$, $\exists\epsilon_\gamma > 0$ such that $\forall\epsilon \in [0, \epsilon_\gamma)$, the GRDE (2.12) admits a positive definite solution, and consequently, the game (2.1)–(2.6) has a finite value. Furthermore, the solution to GRDE (2.12) can be approximated by

$$\bar{Z} = \begin{bmatrix} Z_{s\gamma}(t) + O(\epsilon) & \epsilon(Z'_c(t) + Z'_{cb}(\tau)) + O(\epsilon^2) \\ \epsilon(Z_c(t) + Z_{cb}(\tau)) + O(\epsilon^2) & \epsilon(Z_{f\gamma}(t) + Z_{fb}(\tau)) + O(\epsilon^2) \end{bmatrix}, \quad (4.46)$$

for all $t \in [0, t_f]$, where $Z_{cb}(\tau)$ and $Z_{fb}(\tau)$ are the solutions to (2.22)–(2.23) and as $\tau \to -\infty$ they converge to 0 exponentially in the τ time scale.

(3) $\forall\gamma > \bar{\gamma}$, if we apply the composite controller $\mu^*_{c\gamma}$ to the system, then $\exists\epsilon'_\gamma > 0$ such that

† This result is not included here because the arguments are quite technical, and its counterpart in the finite horizon case still has not been established.

‡ This condition is imposed in order to make the boundary layer term $L_{22}(\tau)$ converge to $Z_{f\gamma}(t_f)$.

$\forall\epsilon \in [0, \epsilon'_\gamma)$, the disturbance attenuation level γ is attained for the full-order system.

(4) $\forall\gamma > \max\{\gamma_s, \gamma_{of}\} = \max\{\hat{\gamma}_s, \gamma_{of}\}$, if we apply only the slow controller $\mu^*_{s\gamma}$ to the system, then $\exists\epsilon''_\gamma > 0$ such that $\forall\epsilon \in [0, \epsilon''_\gamma)$, the disturbance attenuation level γ is attained for the full-order system.

Proof. By a reasoning analogous to that used in the proof of Theorem 1, Assumption A3, together with the controllability of (A_{22}, B_2), leads to finiteness of γ_f, and hence to that of $\bar{\gamma}$. Now, fix $\gamma > \bar{\gamma}$, under Assumptions A1–A3. Then, we get $Z_{f\gamma}(t) > 0$ and that $A_{22}(t) - S_{22}(t)Z_{f\gamma}(t)$ is Hurwitz. Following a reasoning as in the infinite horizon case, we deduce that $Z_{11} = Z_{s\gamma}(t)$, $Z_{12} = Z'_c(t)$ and $Z_{22} = Z_{f\gamma}(t)$ satisfy equations (2.17)–(2.19). We now use Theorem L in Kokotoviĉ and Yackel (1972) to prove 1 and 2. (For convenience to the reader, we include this theorem in Appendix D.)

View equations (2.14)–(2.16) as a nonlinear singularly perturbed system. Obviously, condition L1 of Theorem L is satisfied. Since $Z_{s\gamma}(t)$ is the unique solution for (3.19) (see Başar and Bernhard, 1991) L3 is satisfied. Following some of the reasoning used in the proof of Theorem 1, L4 is established. We only need to show L2, which is equivalent to showing that $L_{22}(\tau)$, $L_{12}(\tau)$ exist, and $L_{22}(\tau) \to Z_{f\gamma}(t_f)$, $L_{12}(\tau) \to Z'_c(t_f)$ as $\tau \to -\infty$, and $Z_{11}(t)(V'_1 + U'_1Z_{f\gamma}(t)) + V'_2 + U'_2Z_{f\gamma}(t)$, $Z_{f\gamma}(t)$ are the asymptotically stable equilibria for (2.20)–(2.21) uniformly in $Z_{11}(t)$ when t_f is replaced by t, where $L_{22}(\tau)$ and $L_{12}(\tau)$ satisfy (2.20)–(2.21). From the proof of Theorem 1, we have that $Z_{11}(t)(V'_1 + U'_1Z_{f\gamma}(t)) + V'_2 + U'_2Z_{f\gamma}(t)$, $Z_{f\gamma}(t)$ are the equilibria for (2.20)–(2.21). Then they are the asymptotically stable equilibria because the linearized system is asymptotically stable.

We now show that $L_{22}(\tau) \to Z_{f\gamma}(t_f)$ $\tau \to -\infty$. If $Q_{f22} = Z_{f\gamma}(t_f)$, the result is trivial. Suppose therefore that $Z_{f\gamma}(t_f) > Q_{f22}$.† Consider (2.21), which we rewrite in the form:

$$\frac{d}{d\tau}L_{22}(\tau)^{-1} - A_{22}(t_f)L_{22}(\tau)^{-1} - L_{22}(\tau)^{-1}A_{22}(t_f)' + S_{22}(t_f) - L_{22}(\tau)^{-1}Q_{22}(t_f)L_{22}(\tau)^{-1} = 0.$$

We evaluate equation (3.46) at $t = t_f$, and rewrite it as:

$$-A_{22}(t_f)Z_{f\gamma}(t_f)^{-1} - Z_{f\gamma}(t_f)^{-1}A_{22}(t_f)' + S_{22}(t_f) - Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1} = 0.$$

Let $\bar{\Delta} := L_{22}(\tau)^{-1} - Z_{f\gamma}(t_f)^{-1}$. Take the difference of the above two equations and obtain,

$$\frac{d}{d\tau}\bar{\Delta} - A_{22}(t_f)\bar{\Delta} - \bar{\Delta}A_{22}(t_f)' - \bar{\Delta}Q_{22}(T_f)\bar{\Delta} - \bar{\Delta}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1} - Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)\bar{\Delta} = 0.$$

Let $\Delta(\tau') := (Z_{f\gamma}(t_f)\bar{\Delta}Z_{f\gamma}(t_f))^{-1}$ where $\tau' = -\tau$. Then the above equation is equivalent to

$$\frac{d}{d\tau'}\Delta + A_f\Delta + \Delta A'_f - Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1} = 0,$$

where $A_f = A_{22}(t_f) - S_{22}(t_f)Z_{f\gamma}(t_f)$. Now introduce the following system:

$$\frac{d}{d\tau'}y = A'_fy; \quad y(0) = y_0.$$

Obviously, this system is stable, and

$$\frac{d}{d\tau'}(y'\Delta y) = y'Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1}y,$$

so that

$$y'_0e^{A_f\tau'}\Delta(\tau')e^{A'_f\tau'}y_0 = y'_0\Big(\Delta(0) + \int_0^{\tau'} e^{A_fs}Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1}e^{A'_fs}\,ds\Big)y_0.$$

Then, since y_0 was arbitrary,

$$\Delta(\tau') = e^{-A_f\tau'}\Big(\Delta(0) + \int_0^{\tau'} e^{A_fs}Z_{f\gamma}(t_f)^{-1}Q_{22}(t_f)Z_{f\gamma}(t_f)^{-1}e^{A'_fs}\,ds\Big)e^{-A'_f\tau'}.$$

Since $Z_{f\gamma}(t_f) > Q_{f22} = L_{22}(0)$, we have $\Delta(0) > 0$, and consequently, $\Delta(\tau') > 0$. Hence $\bar{\Delta}$ exists and is positive definite for every τ. By the stability of $A_{22}(t_f) - S_{22}Z_{f\gamma}(t_f)$, we deduce that $\Delta(\tau') \to +\infty$ as $\tau' \to +\infty$, which implies that $Z_{f\gamma} \to 0$ as $\tau \to -\infty$ and the convergence rate is exponential in the τ time scale.

Now, we rewrite (2.20) as follows:

$$\frac{d}{d\tau}L_{12}(\tau) = -L_{12}(\tau)(A_{22}(t_f) - S_{22}(t_f)Z_{f\gamma}(t_f) + S_{22}(Z_{f\gamma}(t_f) - L_{22}(\tau))) - (A'_{21}(t_f) - Z_{11}(t_f)S_{12}(t_f))L_{22}(\tau) - Z_{11}(t_f)A_{12}(t_f) - Q_{12}(t_f).$$

Since $A_{22}(t_f) - S_{22}(t_f)Z_{f\gamma}(t_f)$ is Hurwitz, and $S_{22}(t_f)(Z_{f\gamma}(t_f) - L_{22}(\tau)) \to 0$ as $\tau \to -\infty$, in view of known results for stability of linear systems (see Coppel, 1965, p. 70, Theorem 9) $L_{12}(\tau) \to$

† It is well known that $L_{22}(\tau)$ converges to $Z_{f\gamma}(t_f)$ as $\tau \to -\infty$ if $Q_{f22} = 0$ (see Başar and Bernhard, 1991) In the cases when Q_{f22} is not strictly positive definite or it is not strictly less than $Z_{f\gamma}$ the results are still true, which follow from the fact that an increase in the terminal condition (in the sense of matrices) leads to an increase in the solution to the RDE.

$Z_c'(t_f)$ exponentially. So L2 is satisfied. Then, by Theorem L, the solution to (2.14)–(2.16) exists for sufficiently small ϵ. Since eigenvalues $\lambda(A_{22}(t)-S_{22}(t)Z_{f\gamma}(t))$ depend continuously on t, there exists a $c>0$ such that $\lambda(A_{22}(t)-S_{22}(t)Z_{f\gamma}(t))\le -c$. Then, by applying Theorem 3.1 in Chapter 1 of Kokotović *et al.* (1986), 1 and 2 follow readily.

For part 3, we first substitute $\mu^*_{c\gamma}$ into system (2.1)–(2.6), and then follow a reasoning similar to that used in the proof of Theorem 1.3 to show (rather easily) that L1, L3, L4 are satisfied. It is only necessary to show L2 to prove 3. Introduce the boundary layer system,

$$\begin{aligned}\frac{d}{d\tau}L_{12}(\tau) &= -L_{12}(\tau)(A_{22}(t_f)-B_2(t_f)B_2'(t_f)Z_{f\gamma}(t_f)\\ &\quad+\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)L_{22}(\tau))-(A_{21}'(t_f)\\ &\quad-Z_{s\gamma}(t_f)S_{12}(t_f)-Z_c'(t_f)B_2(t_f)B_2'(t_f))L_{22}(\tau)\\ &\quad-Z_{s\gamma}(t_f)A_{12}(t_f)-Q_{12}(t_f)-Z_{s\gamma}(t_f)S_{12}(t_f)Z_{f\gamma}(t_f)\\ &\quad-Z_c'(t_f)B_2(t_f)B_2'(t_f)Z_{f\gamma}(t_f),\quad L_{12}(0)=Q_{f12},\end{aligned} \tag{4.47}$$

$$\begin{aligned}\frac{d}{d\tau}L_{22}(\tau) &= -(A_{22}'(t_f)-Z_{f\gamma}(t_f)B_2(t_f)B_2'(t_f))L_{22}(\tau)\\ &\quad-L_{22}(\tau)(A_{22}(t_f)-B_2(t_f)B_2'(t_f)Z_{f\gamma}(t_f))\\ &\quad-Q_{22}(t_f)-Z_{f\gamma}(t_f)B_2(t_f)B_2'(t_f)Z_{f\gamma}(t_f)\\ &\quad-L_{22}(\tau)\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)L_{22}(\tau);\\ L_{22}(0)&=Q_{f22}\end{aligned} \tag{4.48}$$

It is obvious that $Z_{11}(t)(V_1'+U_1'Z_{f\gamma}(t))+V_2'+U_2'Z_{f\gamma}(t)$, $Z_{f\gamma}(t)$ are the equilibria for the above system with t_f substituted by t. Then they are the asymptotically stable equilibria because the linearized system is asymptotically stable. Now consider (4.48). Since the result is trivially true when $Q_{f22}=Z_{f\gamma}(t_f)$, we only consider the case $Z_{f\gamma}(t_f)>Q_{f22}$. Let A_f be defined as before and $\bar\Delta(\tau):=Z_{f\gamma}(t_f)-L_{22}(\tau)$. Subtract (4.48) from (3.46) at $t=t_f$, to arrive at

$$\frac{d}{d\tau}\bar\Delta+A_f'\bar\Delta+\bar\Delta A_f-\bar\Delta\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)\bar\Delta=0.$$

Let $\Delta(\tau')=\bar\Delta(\tau)^{-1}$ where $\tau'=-\tau$. Then

$$\frac{d}{d\tau'}\Delta+A_f\Delta+\Delta A_f'-\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)=0.$$

Define the corresponding system:

$$\frac{d}{d\tau'}y=A_f'y;\quad y(0)=y_0.$$

Obviously, this system is stable and

$$\frac{d}{d\tau'}(y'\Delta y)=y'\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)y,$$

so that

$$\begin{aligned}y_0'e^{A_f\tau'}\Delta(\tau')e^{A_f'\tau'}y_0 &= y_0'\Big(\Delta(0)\\ &\quad+\int_0^{\tau'}e^{A_f s}\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)e^{A_f' s}\,ds\Big)y_0.\end{aligned}$$

Then, since y_0 was arbitrary,

$$\begin{aligned}\Delta(\tau') &= e^{-A_f\tau'}\Big(\Delta(0)+\\ &\quad+\int_0^{\tau'}\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)e^{A_f' s}\,ds\Big)e^{-A_f'\tau'}\end{aligned}$$

Since $Z_{f\gamma}(t_f)>Q_{f22}=L_{22}(0)$, we have $\Delta(0)>0$, and consequently, $\Delta(\tau')>0$. Hence, $\bar\Delta$ exists and is positive definite for every τ. By the stability of $A_{22}(t_f)-S_{22}Z_{f\gamma}(t_f)$, we deduce that $\Delta(\tau')\to+\infty$ as $\tau'\to+\infty$, which implies that $L_{22}(\tau)\to Z_{f\gamma}(t_f)$ as $\tau\to-\infty$. Then, by an argument similar to that used in 1 and 2, $L_{12}(\tau)\to Z_c'(t_f)$ as $\tau\to-\infty$. Thus, L2 is established. Then 3 follows from Theorem L.

To prove part 4, we use a similar procedure. We first substitute $\mu^*_{s\gamma}$ into the system (2.1)–(2.6), and follow the reasoning of Theorem 1(4), to show that L1, L3 and L4 are satisfied. To show L2, we introduce the boundary layer system,

$$\begin{aligned}\frac{d}{d\tau}L_{12}(\tau) &= -L_{12}(\tau)(A_{22}(t_f)\\ &\quad+\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)L_{22}(\tau))\\ &\quad-(A_{21}'(t_f)-Z_{s\gamma}(t_f)S_{12}(t_f)\\ &\quad-V'(t_f)B_2(t_f)B_2'(t_f))L_{22}(\tau)\\ &\quad-Z_{s\gamma}(t_f)A_{12}(t_f)-Q_{12}(t_f),\\ L_{12}(0)&=Q_{f12},\end{aligned} \tag{4.49}$$

$$\begin{aligned}\frac{d}{d\tau}L_{22}(\tau) &= -A_{22}'(t_f)L_{22}(\tau)-L_{22}(\tau)A_{22}(t_f)\\ &\quad-Q_{22}(t_f)+L_{22}(\tau)\frac{1}{\gamma^2}D_2(t_f)D_2'(t_f)L_{22}(\tau),\\ L_{22}(0)&=Q_{f22}.\end{aligned} \tag{4.50}$$

Since $Z_{of\gamma}(t_f)\ge Z_{f\gamma}(t_f)\ge Q_{f22}$, we can use a reasoning similar to that used in the proof of 3 to show that $L_{22}(\tau)\to Z_{of\gamma}(t_f)$ as $\tau\to-\infty$, and subsequently $L_{12}(\tau)\to V'+U'Z_{of\gamma}(t_f)$ as $\tau\to-\infty$. Thus, L2 is satisfied, which then establishes 4.

5 EXAMPLES

We present in this section some numerical results for the infinite horizon case. As stressed earlier, the three quantities $\gamma_{s\infty}$, $\gamma_{f\infty}$ and $\gamma_{of\infty}$ play

important roles in the computation of an approximate value for $\gamma_{\infty}^*(\epsilon)$. Furthermore, even though we know the order relationship $\gamma_{\text{of}\infty} \geq \gamma_{f\infty}$, there is no such relationship between $\gamma_{s\infty}$ and the other two quantities, and it is very much problem dependent as we will see shortly. In the examples below, we will also study the performances attained by the approximate controllers $\mu_{c\gamma}^*$, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$, when they are applied to the original system.†

Example 1.

Consider the following two-dimensional system and the performance index:

$$\begin{bmatrix} \dot{x}_1 \\ \epsilon \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} u + \begin{bmatrix} 1 \\ 3 \end{bmatrix} w, \quad (5.1)$$

$$L_\gamma = \int_0^{t_f} (2x_1^2 + 2x_1x_2 + x_2^2 + |u|^2 - \gamma^2 |w|^2)\, dt. \quad (5.2)$$

Note that here the fast system is open-loop unstable. By using a particular search algorithm, we readily compute the three basic quantities:

$$\gamma_{s\infty} = 2.3565,$$

$$\gamma_{f\infty} = 3.0000,$$

$$\gamma_{\text{of}\infty} = \infty$$

Note that here the three values satisfy the relationship $\gamma_{s\infty} < \gamma_{f\infty} < \gamma_{\text{of}\infty}$. We can also compute the minimax disturbance attenuation level $\gamma_{\infty}^*(\epsilon)$ of the full-order system (5.1)–(5.2) for different values of ϵ as shown in Table 1. Note that as $\epsilon \to 0$, $\gamma^*(\epsilon) \to \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$.

TABLE 1. OPTIMUM H^∞-PERFORMANCE LEVEL FOR THE FULL ORDER SYSTEM OF EXAMPLE 1

$\gamma^*(\epsilon)$	2.4332	2.7843	2.9724	2.9972	2.9997
ϵ	1	0.1	0.01	0.001	0.0001

Now, we choose $\gamma = 3.5 > \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$, and design the optimal controller for the slow and fast subsystems based on this value of γ:

$$Z_{s\gamma} = 0.4822,$$

$$\mu_{s\gamma}^*(x_1) = -0.31378x_1,$$

$$\mu_{sf\gamma}^*(x_1, w) = 0.38576x_1 - 0.60000w,$$

$$v_{s\gamma}^*(x_1) = -0.11997x_1,$$

$$Z_{f\gamma} = 15.3229,$$

$$\mu_{f\gamma}^*(x_1, x_2) = -10.1614x_1 - 15.3229x_1,$$

$$\mu_{c\gamma}^*(x_1, x_2) = -10.4752x_1 - 15.3229x_2.$$

† In these examples, the relative accuracy is 0.002

Then, we apply $\mu_{c\gamma}^*$, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$ to the full-order system (5.1)–(5.2) and obtain the corresponding disturbance attenuation bounds γ_c^*, γ_s^* and γ_{sf}^*, respectively, which are tabulated in Table 2.

TABLE 2. A COMPARISON OF H^∞-PERFORMANCE LEVELS UNDER COMPOSITE, SLOW AND FEEDFORWARD CONTROLLERS FOR EXAMPLE 1

γ_c^*	3.4818	3.4818	3.4818	3.4818	3.4818
γ_s^*	∞	∞	∞	∞	∞
γ_{sf}^*	∞	∞	∞	∞	∞
ϵ	1	0.1	0.01	0.001	0.0001

Note that only the composite controller $\mu_{c\gamma}^*$ achieves the desired γ bound level. Even if one allows a disturbance feedforward design based on the slow subsystem only (which is the controller $\mu_{sf\gamma}^*$), still the fast dynamics cannot be stabilized.

Example 2.

Consider the following system and the performance index:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \end{bmatrix} u + \begin{bmatrix} 1 \\ 3 \end{bmatrix} w, \quad (5.3)$$

$$L_\gamma = \int_0^{t_f} (2x_1^2 + 2x_1x_2 + 3x_2^2 + |u|^2 - \gamma^2 |w|^2)\, dt. \quad (5.4)$$

Here, the fast subsystem is open-loop stable. The three basic performance levels in this case are:

$$\gamma_{s\infty} = 1.4240,$$

$$\gamma_{f\infty} = 1.2990,$$

$$\gamma_{\text{of}\infty} = 2.5981.$$

Note that here the three values satisfy the relationship $\gamma_{f\infty} < \gamma_{s\infty} < \gamma_{\text{of}\infty}$. We can also compute the minimax disturbance attenuation level $\gamma_{\infty}^*(\epsilon)$ of the system (5.3)–(5.4) for every fixed ϵ (see Table 3). Note again that as $\epsilon \to 0$, $\gamma^*(\epsilon) \to \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$. Now, we choose $\gamma = 1.6 > \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$, and design the optimal controller for the slow and fast subsystems based on this value of γ:

$$Z_{s\gamma} = 3.3788,$$

$$\mu_{s\gamma}^*(x_1) = 1.6447x_1,$$

$$\mu_{sf\gamma}^*(x_1, w) = -2.4091x_1 - 1.1250w,$$

$$v_{s\gamma}^*(x_1) = -3.6034x_1,$$

$$Z_{f\gamma} = 0.69201,$$

$$\mu_{f\gamma}^*(x_1, x_2) = -5.88965x_1 - 1.3840x_2,$$

$$\mu_{c\gamma}^*(x_1, x_2) = -4.22518x_1 - 1.3840x_2.$$

TABLE 3 OPTIMUM H^∞-PERFORMANCE LEVEL FOR THE FULL ORDER SYSTEM OF EXAMPLE 2

$\gamma^*(\epsilon)$	1 3306	1 3993	1 4212	1 4237	1 4240
ϵ	1	0 1	0 01	0 001	0 0001

TABLE 4 A COMPARISON OF H^∞ PERFORMANCE LEVELS UNDER COMPOSITE SLOW AND FEEDFORWARD CONTROLLERS FOR EXAMPLE 2 WITH $\gamma = 1\ 6$

γ_c^*	1 4497	1 4497	1 4497	1 4497	1 4497
γ_s^*	∞	∞	∞	∞	∞
γ_{sf}^*	1 5439	1 5439	1 5439	1 5439	1 5439
ϵ	1	0 1	0 01	0 001	0 0001

Then, we use $\mu_{c\gamma}^*$, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$ in system (5 1)–(5 2) and obtain corresponding disturbance attenuation bounds γ_c^*, γ_s^* and γ_{sf}^*, respectively, which are tabulated in Table 4 It is clear that in this case both the composite and the feedforward controllers, $\mu_{c\gamma}^*$ and $\mu_{sf\gamma}^*$, can achieve the γ bound level, but the slow controller leads to infinite attenuation Also, the level achieved by feedforward controller is above the level achieved by $\mu_{c\gamma}^*$ If we want to totally ignore the fast dynamics in the design of the controller then we should choose γ to be greater than the $\gamma_{of\infty}$ in the design of the controllers For instance, if we choose $\gamma = 2\ 7$, we obtain the value

$$Z_{s\gamma} = 2\ 1713$$
$$\mu_{s\gamma}^*(x_1) = -1\ 0843x_1,$$
$$\mu_{sf\gamma}^*(x_1, w) = -1\ 5034x_1 - 1\ 125w$$
$$v_{s\gamma}^*(x_1) = -0\ 37258x_1,$$
$$Z_{f\gamma} = 0\ 54480,$$
$$\mu_{f\gamma}^*(x_1, x_2) = \ 2\ 3352x_1 - 1\ 0896x_2,$$
$$\mu_{c\gamma}^*(x_1, x_2) = -3\ 4195x_1 - 1\ 0896x_2$$

Then, using $\mu_{c\gamma}^*$, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$ in the system (5 3)–(5 4), we obtain the corresponding disturbance attenuation bounds γ_c^*, γ_s^* and γ_{sf}^*, respectively, as shown in Table 5 Note that in this case all three controllers achieve the 2 7 bound, but obviously, the composite controller $\mu_{c\gamma}^*$ does much better than the other two controllers, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$, not only for small values of $\epsilon > 0$, but even for $\epsilon = 1$ in which case $\mu_{s\gamma}^*$ does not yield the desired bound (but $\mu_{sf\gamma}^*$ does)

TABLE 5 A COMPARISON OF H^∞ PERFORMANCE LEVELS UNDER COMPOSITE SLOW AND FEEDFORWARD CONTROLLERS FOR EXAMPLE 2 WITH $\gamma = 2\ 7$

γ_c^*	1 3899	1 4152	1 4344	1 4533	1 4533
γ_s^*	3 5891	2 7145	2 5315	2 5990	2 5990
γ_{sf}^*	1 7093	1 6838	1 6838	1 6838	1 6838
ϵ	1	1	0 01	0 001	0 0001

Example 3

Now consider the following system and performance index

$$\begin{bmatrix} \dot{x}_1 \\ \epsilon \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}u + \begin{bmatrix} 1 \\ 1 \end{bmatrix}w, \quad (5\ 5)$$

$$L_\gamma = \int_0^\infty (2x_1^2 + 2x_1x_2 + x_2^2 + |u|^2 - \gamma^2 |w|^2)\, dt \quad (5\ 6)$$

We again compute the three basic quantities

$$\gamma_{s\infty} = 0\ 63249,$$
$$\gamma_{f\infty} = 0\ 44697,$$
$$\gamma_{of\infty} = 0\ 49941$$

In this case, these satisfy the relationship $\gamma_{f\infty} < \gamma_{of\infty} < \gamma_{s\infty}$ We also compute $\gamma_\epsilon^*(\epsilon)$ for different values of $\epsilon > 0$ (see Table 6) Again, as $\epsilon \to 0$, $\gamma^*(\epsilon) \to \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$ Now, we choose $\gamma = 0\ 7 > \max\{\gamma_{s\infty}, \gamma_{f\infty}\}$, and design the optimal controller for the slow and fast subsystems based on this value of γ

TABLE 6 OPTIMUM H^∞ PERFORMANCE LEVEL FOR THE FULL ORDER SYSTEM OF EXAMPLE 3

$\gamma^*(\epsilon)$	0 55946	0 61434	0 63023	0 63223	0 63243
ϵ	1	0 1	0 01	0 001	0 0001

$$Z_{s\gamma} = 5\ 5839,$$
$$\mu_{s\gamma}^*(x_1) = -16\ 370x_1,$$
$$\mu_{sf\gamma}^*(x_1, w) = -11\ 968x_1 - 0\ 20{,}000w,$$
$$v_{s\gamma}^*(x_1) = 22\ 013x_1,$$
$$Z_{f\gamma} = 0\ 26880,$$
$$\mu_{f\gamma}^*(x_1, x_2) = 1\ 0272x_1 - 0\ 26880x_2,$$
$$\mu_{c\gamma}^*(x_1, x_2) = -15\ 343x_1 - 0\ 26880x_2$$

Using $\mu_{c\gamma}^*$, $\mu_{s\gamma}^*$ and $\mu_{sf\gamma}^*$ in systems (5 5)–(5 6), obtain the values shown in Table 7 for the corresponding disturbance attenuation bounds

TABLE 7 A COMPARISON OF H^∞-PERFORMANCE LEVELS UNDER COMPOSITE SLOW AND FEEDFORWARD CONTROLLERS FOR EXAMPLE 3

γ_c^*	0 69507	0 69507	0 69507	0 69507	0 69507
γ_s^*	0 69713	0 69713	0 69713	0 69713	0 69713
γ_{sf}^*	0 69309	0 69309	0 69309	0 69309	0 69309
ϵ	1	0 1	0 01	0 001	0 0001

γ_c^*, γ_s^* and γ_{sf}^*, respectively. We can see that all three controllers achieve the 0.7 bound, but $\mu_{sf\gamma}^*$ does a little better than the other two. This is mainly due to the fact that in this case $\gamma_{of\infty} < \gamma_{s\infty}$.

We observe from the preceding analysis the following essential features of approximate designs based on time scale separation:

(1) When the fast subsystem is open-loop unstable, only the composite controller $\mu_{c\gamma}^*$ can achieve a $\gamma > \bar{\gamma}_\infty$ bound.

(2) Even if the fast subsystem is open-loop stable, the composite controller may achieve a better disturbance attenuation bound than the $\mu_{sf\gamma}^*$ when $\gamma_{s\infty} < \gamma_{of\infty}$.

(3) In the case of $\gamma_{s\infty} \geq \gamma_{of_\infty} \geq \gamma_{f_\infty}$, all three controllers (i.e. $\mu_{s\gamma}^*$, $\mu_{sf\gamma}^*$ and $\mu_{c\gamma}^*$) can achieve a $\gamma > \bar{\gamma}_\infty$. The composite controller does not necessarily lead to a much better performance than the other two.

6. CONCLUSION

In this paper, we have provided a complete analysis of the singularly perturbed linear-quadratic H^∞-optimal control problem, with perfect state measurements in both finite and infinite horizons, by relating it to a class of singularly perturbed differential games. One of the main results of the paper is the existence and construction of a composite controller, independent of the singular perturbation parameter, under which the associated differential game has a bounded upper value. This composite controller leads to a specific (pre-computable) H^∞-performance bound for the full-order control problem for sufficiently small values of the singular perturbation parameter. For the infinite-horizon case, we have actually proved recently, in Pan and Başar (1992), that this performance bound is tight;† we have not included this result here since the arguments are quite technical and lengthy, and its counterpart in the finite horizon case is still an open problem.

In the present paper, we have also obtained the conditions under which a controller design based on the slow subsystem only can achieve a desired performance level. Yet another controller whose performance is studied is the one that uses slow state feedback and a feedforward term from the disturbance, and it has been shown that the performance of this controller is generally inferior to that of the composite controller. One of the important messages of the paper has been that the "best" composite controller is not simply the sum of the "best" slow and the "best" fast controllers (contrary to what is observed in the singularly perturbed LQR problem) but involves a more intricate construction.

One immediate, but not trivial, extension of these results would be to the sampled-data measurement case, so as to obtain the counterparts of the results of Başar (1991) in the singularly perturbed case. This work has already been completed and the results will be presented elsewhere (Pan and Başar (1991b)). Another extension is to the imperfect measurement case (the four-block H^∞-optimal control problem) which has also recently been completed by Pan and Başar (1991a). One other extension would be to the multiple-time scale problems, which is a topic currently under study.

† In the recent work, Pan and Başar (1992), we prove a weak converse of Theorem 2, which says that $\lim_{\epsilon\to 0+} \gamma^*(\epsilon) = \bar{\gamma}$ provided $\gamma_s > \gamma_f$ and the matrix $Q_f > 0$ for small enough, but positive, ϵ.

REFERENCES

Başar, T. (1991). Optimum H^∞ designs under sampled state measurements. *Systems and Control Letters*, 399–409.

Başar, T. and P. Bernhard (1991). *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach* Birkhäuser, Boston.

Başar, T. and G. J. Olsder (1982). *Dynamic Noncooperative Game Theory*. Academic Press, London.

Chow, J. H. and P. V. Kokotović (1976) A decomposition of near-optimum regulators for systems with slow and fast modes *IEEE Trans. on Aut. Control*, **AC-21**, 701–705

Coppel, W. A. (1965) *Stability and Asymptotic Behavior of Differential Equations* D.C. Heath and Company, Boston

Corless, M., G. Leitmann and E. P. Ryan (1990). Control of uncertain systems with neglected dynamics. In A. S. I. Zinober (Ed.), *Deterministic Control of Uncertain Systems*, pp. 252–268, IEE Press, London

Gardner, Jr., B. F. and J. B. Cruz, Jr. (1978). Well-posedness of singularly perturbed Nash games. *J. of the Franklin Institute*, **306**, 355–374.

Garofalo, F. and G. Leitmann (1988). A composite controller ensuring ultimate boundedness for a class of singularly perturbed uncertain systems. *Dynamics and Stability of Systems*, **3**, 135–145

Garofalo, F. and G. Leitmann (1990). Nonlinear composite control of a class of nominally linear singularly perturbed uncertain systems In A. S. I. Zinober (Ed.), *Deterministic Control of Uncertain Systems*, pp. 269–288, IEE Press, London.

Kokotović, P. V. and R. A. Yackel (1972). Singular perturbation of linear regulators: Basic theorems *IEEE Trans. on Aut. Control*, **AC-17**, pp. 29–37.

Kokotović, P. V., H. K. Khalil and J. O'Reilly (1986). *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press, New York

Luse, D. W. and J. A. Ball (1989). Frequency-scale decomposition of H^∞-disk problems. *SIAM J. Control and Optimization*, **27**, 814–835.

Pan, Z. and T. Başar (1991a). H^∞-optimal control for singularly perturbed systems. Part II: Imperfect state measurements. CSL Report, University of Illinois, Urbana, IL. November; *Proc. 31st IEEE Conf. on Decision and Control*, Tucson, Arizona, 16–18 Dec. 1992 (to appear).

Pan, Z. and T. Başar (1991b). H^∞-optimal control of singularly perturbed systems with Sampled-state measurements. CSL Report, University of Illinois, Urbana, IL. November; to be presented at the *Fifth International Symposium on Dynamic Games and Applications*, Geneve, Switzerland, July 1992.

Pan, Z. and T. Başar (1992). A tight bound for the H^∞ performance of singularly perturbed systems. CSL Report, University of Illinois, Urbana, IL.

APPENDIX A SIMPLIFICATION OF SUBSYSTEM PARAMETER MATRICES

First, we introduce some matrices to simplify our proof

$$\pi := A_{22}Q_{22}^{-1}A'_{22}, \quad (A1)$$

$$\rho := I + B'_2A'^{-1}_{22}Q_{22}A_{22}^{-1}B_2, \quad (A2)$$

$$\sigma = \gamma^2 I - D'_2(A_{22}Q_{22}^{-1}A'_{22} + B_2B'_2)^{-1}D_2, \quad (A3)$$

$$\phi = \gamma^2 I - D'_2A'^{-1}_{22}Q_{22}A_{22}^{-1}D_2, \quad (A4)$$

$$\chi = I + B'_2\Big(A_{22}Q_{22}^{-1}A'_{22} - \frac{1}{\gamma^2}D_2D'_2\Big)^{-1}B_2 \quad (A5)$$

In terms of these matrices, we can obtain the following expressions using some simple matrix operations and matrix inversion identities

$$\rho = I + B'_2\pi^{-1}B_2, \quad (A6)$$

$$\rho^{-1} = I - B'_2(\pi + B_2B'_2)^{-1}B_2, \quad (A7)$$

$$\rho^{-1}B'_2\pi^{-1} = B'_2(\pi + B_2B'_2)^{-1}, \quad (A8)$$

$$\sigma = \gamma^2 I - D'_2(\pi + B_2B'_2)^{-1}D_2, \quad (A9)$$

$$\sigma^{-1} = \frac{1}{\gamma^2}I + \frac{1}{\gamma^4}D'_2(\pi + S_{22})^{-1}D_2, \quad (A10)$$

$$\sigma^{-1}D'_2(\pi + B_2B'_2)^{-1} = \gamma^2 D'_2(\pi + S_{22})^{-1}, \quad (A11)$$

$$\phi = \gamma^2 I - D'_2\pi^{-1}D_2, \quad (A12)$$

$$\phi^{-1} = \frac{1}{\gamma^2}\Big(I + \frac{1}{\gamma^2}D'_2\Big(\pi - \frac{1}{\gamma^2}D_2D'_2\Big)^{-1}D_2\Big), \quad (A13)$$

$$\phi^{-1}D'_2\pi^{-1} = \frac{1}{\gamma^2}D'_2\Big(\pi - \frac{1}{\gamma^2}D_2D'_2\Big)^{-1} \quad (A14)$$

$$\chi^{-1} = I - B'_2(\pi + S_{22})^{-1}B_2, \quad (A15)$$

$$\chi^{-1}B'_2\Big(\pi - \frac{1}{\gamma^2}D_2D'_2\Big)^{-1} = B'_2(\pi + S_{22})^{-1} \quad (A16)$$

Transformations of the slow subsystem under the H information pattern

We will deal with the transformations described in (3.8) and (3.9) one by one. Substitution of (3.8) into (3.2) and (3.7) yields the following system

$$\dot{x}_s = A^{\#}x_s + B^{\#}u_s + D^{\#}w_s, \quad x_s(0) = x_1(0), \quad (A17)$$

$$L_{\gamma s} = |x_s(t_f)|^2_{Q_{f11}} + \int_0^{t_f}(|x_s|^2_{Q^{\#}} + 2x'_sP^{\#}w_s + |w_s|^2_{R^{\#}} + |\bar{u}_s|^2)\,dt. \quad (A18)$$

where

$$\begin{aligned}
A^{\#} &= A_0 + B_0\rho^{-1}B'_2\pi^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= A_0 + B_0B'_2(\pi + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21}),\\
B^{\#} &= B_0\rho^{-1/2},\\
D^{\#} &= D_0 - B_0\rho^{-1}B'_2\pi^{-1}D_2\\
&= D_0 - B_0B'_2(\pi + B_2B'_2)^{-1}D_2,\\
Q^{\#} &= Q_{11} - A'_{21}A'^{-1}_{22}Q_{21} - Q_{12}A_{22}^{-1}A_{21}\\
&\quad + A'_{21}\pi^{-1}A_{21} - (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})\pi^{-1}\\
&\quad \times B_2\rho^{-1}B'_2\pi^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= Q_{11} - A'_{21}A'^{-1}_{22}Q_{21} - Q_{12}A_{22}^{-1}A_{21} + A'_{21}\pi^{-1}A_{21}\\
&\quad - (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})\pi^{-1}B_2B'_2(\pi\\
&\quad + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})\\
&\quad \times (\pi + B_2B'_2)^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}),\\
P^{\#} &= -Q_{12}A_{22}^{-1}D_2 + A'_{21}\pi^{-1}D_2\\
&\quad + (Q_{12}D_{22}^{-1}A'_{22} - A'_{21})\pi^{-1}B_2\rho^{-1}B'_2\pi^{-1}D_2\\
&= -Q_{12}A_{22}^{-1}D_2 + A'_{21}\pi^{-1}D_2\\
&\quad + Q_{12}Q_{22}^{-1}A'22 - A'_{21})\\
&\quad \times (\pi + B_2B'_2)^{-1}B_2B'_2\pi^{-1}D_2\\
&= -(Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})(\pi + B_2B'2)^{-1}D_2,\\
R^{\#} &= -\gamma^2 I + D'_2\pi^{-1}D_2 - D'_2\pi^{-1}B_2\rho^{-1}B'_2\pi^{-1}D_2\\
&= -\gamma^2 I + D'_2\pi^{-1}D_2 - D'_2\pi^{-1}B_2B'_2\\
&\quad \times (\pi + B_2B'_2)^{-1}D_2\\
&= -\gamma^2 I + D'_2(\pi + B_2B'_2)^{-1}D_2\\
&= -\sigma
\end{aligned}$$

Now, substitute (3.9) into (A17) and (A18) to obtain

$$\begin{aligned}
\bar{A}_0 &= A_0 + B_0B'_2(\pi + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21}\\
&\quad - A_{21}) - (D_0 - B_0B'_2(\pi + B_2B'_2)^{-1}D_2)\\
&\quad \times \sigma^{-1}D'_2(\pi + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= A_0 + B_0B'_2(\pi + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21}\\
&\quad - A_{21}) - (D_0 - B_0B'_2(\pi + B_2B'_2)^{-1}D_2)\\
&\quad \times \frac{1}{\gamma^2}D'_2(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= A_0 + B_0B'_2(\pi + B_2B'_2)^{-1}\Big(I + \frac{1}{\gamma^2}D_2D'_2(\pi + S_{22})^{-1}\Big)\\
&\quad \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}) - \frac{1}{\gamma^2}D_0D'_2(\pi\\
&\quad + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= A_0 + (S_{12} - A_{12}A_{22}^{-1}S_{22})(\pi + S_{22})^{-1}\\
&\quad \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= A_{11} - A_{12}Q_{22}^{-1}Q_{21} + (S_{12} + A_{12}Q_{22}^{-1}A'_{22})\\
&\quad \times (\pi + S_{22})^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}),\\
\bar{B}_0 &= B^{\#},\\
D_0 &= D^{\#}\sigma^{-1/2},\\
\bar{Q} &= Q^{\#} + (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})(\pi\\
&\quad + B_2B'_2)^{-1}D_2\sigma^{-1}D'_2\\
&\quad \times (\pi + B_2B'_2)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21}),\\
&= Q^{\#} + (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})(\pi + B_2B'_2)^{-1}\frac{1}{\gamma^2}D_2D'_2\\
&\quad \times (\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\\
&= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})(\pi + S_{22})^{-1}\\
&\quad \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})
\end{aligned}$$

Now, we can compute S_0 as follows,

$$\begin{aligned}
S_0 &= \bar{B}_0\bar{B}'_0 - \bar{D}_0\bar{D}'_0\\
&= B_0B'_0 - B_0B'_2(\pi + B_2B'_2)^{-1}B_2B'_0 - D_0\sigma^{-1}D'_0\\
&\quad + B_0B'_2(\pi + S_{22})^{-1}\frac{1}{\gamma^2}D_2D'_0 + \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}B_2B'_0\\
&\quad - B_0B'_2(\pi + B_2B'_2)^{-1}\frac{1}{\gamma^2}D_2D'_2(\pi + S_{22})^{-1}B_2B'_0\\
&= B_0B'_0 - B_0B'_2(\pi + S_{22})^{-1}B_2B'_0\\
&\quad - \frac{1}{\gamma^2}D_0D'_0 - \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}\frac{1}{\gamma^2}D_2D'_0\\
&\quad + B_0B'_2(\pi + S_{22})^{-1}\frac{1}{\gamma^2}D_2D'_0 + \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}B_2B'_0\\
&= S_{11} - A_{12}A_{22}^{-1}S_{21} - S_{12}A'^{-1}_{22}A'_{12} + A_{12}A_{22}^{-1}S_{22}A'^{-1}_{22}A'_{12}\\
&\quad - (S_{12} - A_{12}A_{22}^{-1}S_{22})(\pi + S_{22})^{-1}(S_{21} - S_{22}A'^{-1}_{22}A'_{12})\\
&= S_{11} + A_{12}Q_{22}^{-1}A'_{12} - (S_{12} + A_{12}Q_{22}^{-1}A'_{22})\\
&\quad \times (\pi + S_{22})^{-1}(S_{21} + A_{22}Q_{22}^{-1}A'_{22})
\end{aligned}$$

Below, we now provide the details of the steps that led

from (3.24)–(3.25) to (3.26)–(3.28).

$$w^*_{s\gamma} = \sigma^{-1}[D^{\#}Z_{s\gamma} - D_2'(\pi + B_2B_2')^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})]x_s$$

$$= \Big[\frac{1}{\gamma^2}D_0'Z_{s\gamma} + \frac{1}{\gamma^2}D_2'(\pi + S_{22})^{-1}D_2D_0'Z_{s\gamma} - \frac{1}{\gamma^2} \times D_2'(\pi + S_{22})^{-1}B_2B_0'Z_{s\gamma} - \frac{1}{\gamma^2}D_2'(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\Big]x_s$$

$$= \Big[\frac{1}{\gamma^2}D_1'Z_{s\gamma} - \frac{1}{\gamma^2}D_2'A_{12}'Z_{s\gamma} - \frac{1}{\gamma^2}D_2' \times (\pi + S_{22})^{-1}(S_{21} - A_{12}A_{22}^{-1}S_{22})Z_{s\gamma} - \frac{1}{\gamma^2}D_2'(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})\Big]x_s$$

$$= \Big[\frac{1}{\gamma^2}D_1'Z_{s\gamma} - \frac{1}{\gamma^2}D_2'(\pi + S_{22})^{-1} \times ((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} - (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))\Big]x_s,$$

$$u^*_{sf\gamma} = -\rho^{-1}[B_0'Z_{s\gamma}x_s + B_2'\pi^{-1}(D_2w_s - (A_{22}Q_{22}^{-1}Q_{21} - A_{21})x_s)]$$

$$= -(B_1'Z_{s\gamma}x_s - B_2'(\pi + B_2B_2')^{-1}B_2B_1'Z_{s\gamma}x_s + B_2'(\pi + B_2B_2')^{-1} \times (D_2w_s - (A_{22}Q_{22}^{-1}Q_{21} - A_{21})x_s - A_{22}Q_{22}^{-1}A_{12}'Z_{s\gamma}x_s))$$

$$= -B_1'Z_{s\gamma}x_s + B_2'(\pi + B_2B_2')^{-1}((B_2B_1' + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} + (A_{22}Q_{22}^{-1}Q_{21} - A_{21}))x_s - B_2'(\pi + B_2B_2')^{-1}D_2w_s,$$

$$u^*_{s\gamma} = -B_1'Z_{s\gamma}x_s + B_2'(\pi + B_2B_2')^{-1}((B_2B_1' + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} + (A_{22}Q_{22}^{-1}Q_{21} - A_{21}))x_s - B_2'(\pi + B_2B_2')^{-1}D_2\Big[\frac{1}{\gamma^2}D_1'Z_{s\gamma} - \frac{1}{\gamma^2}D_2'(\pi + S_{22})^{-1}((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} - (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))\Big]x_s$$

$$= -B_1'Z_{s\gamma}x_s + B_2'(\pi + B_2B_2')^{-1}((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} + (A_{22}Q_{22}^{-1}Q_{21} - A_{21}))x_s + B_2'(\pi + B_2B_2')^{-1}\frac{1}{\gamma^2}D_2D_2'(\pi + S_{22})^{-1} \times ((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} - (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))x_s$$

$$= -B_1'Z_{s\gamma}x_s + B_2'(\pi + S_{22})^{-1}((S_{21} + A_{22}Q_{22}^{-1}A_{12}')Z_{s\gamma} - (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))x_s$$

Transformations of the slow subsystem under the SF information pattern

Similar to that in the FI information case, we consider the transformations introduced in (3.29) and (3.30) one at a time. Substitution of (3.29) into (3.2) and (3.7) leads to the following system

$$\dot{x}_s = A^{\square}x_s + B^{\square}u_s + D^{\square}\bar{w}_s, \quad x_s(0) = x_1(0), \tag{A19}$$

$$J_{s\gamma} = |x_s(t_f)|^2_{Q_{f11}} + \int_0^{t_f}(|x_s|^2_{Q^{\square}} + 2x_s'P^{\square}u_s + |u_s|^2_{R^{\square}} - |\bar{w}_s|^2)\,dt, \tag{A20}$$

where

$$A^{\square} = A_0 + D_0\phi^{-1}D_2'\pi^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21}) = A_0 - \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}),$$

$$B^{\square} = B_0 + D_0\phi^{-1}D_2'\pi^{-1}B_2 = B_0 + \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2,$$

$$D^{\square} = D_0\phi^{-1/2},$$

$$Q^{\square} = Q_{11} - A_{21}'A_{22}'^{-1}Q_{21} - Q_{12}A_{22}^{-1}A_{21} + A_{21}'\pi^{-1}A_{21} + (Q_{12}Q_{22}^{-1}A_{22}' - A_{21}')\pi^{-1}D_2\phi^{-1}D_2'\pi^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= Q_{11} - A_{21}'A_{22}'^{-1}Q_{21} - Q_{12}A_{22}^{-1}A_{21} + A_{21}'\pi^{-1}A_{21} + (Q_{12}Q_{22}^{-1}A_{22}' - A_{21}')\pi^{-1}\frac{1}{\gamma^2}D_2D_2' \times \Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (Q_{12}Q_{22}^{-1}A_{22}' - A_{21}') \times \Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21}),$$

$$P^{\square} = -Q_{12}A_{22}^{-1}B_2 + A_{21}'\pi^{-1}B_2 - (Q_{12}Q_{22}^{-1}A_{22}' - A_{21}')\pi^{-1}D_2\phi^{-1}D_2'\pi^{-1}B_2$$

$$= -Q_{12}A_{22}^{-1}B_2 + A_{21}'\pi^{-1}B_2 - (Q_{12}Q_{22}^{-1}A_{22}' - A_{21}')\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}\frac{1}{\gamma^2}D_2D_2'\pi^{-1}B_2$$

$$= -(Q_{12}Q_{22}^{-1}A_{22}' - A_{21}')\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2,$$

$$R^{\square} = I + B_2'\pi^{-1}B_2 + B_2'\pi^{-1}D_2\phi^{-1}D_2'\pi^{-1}B_2$$

$$= I + B_2'\pi^{-1}B_2 + B_2'\pi^{-1}\frac{1}{\gamma^2}D_2D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2$$

$$= I + B_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2$$

$$= \chi.$$

Now, substitute (3.30) into (A19) and (A20) to obtain

$$\hat{A}_0 = A_0 - \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}) + \Big(B_0 + \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2\Big)\chi^{-1}B_2' \times \Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= A_0 - \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}) + \Big(B_0 + \frac{1}{\gamma^2}D_0D_2'\Big(\pi - \frac{1}{\gamma^2}D_2D_2'\Big)^{-1}B_2\Big)B_2'(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= A_0 + B_0B_2'(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21}) - \frac{1}{\gamma^2}D_0D_2'(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= A_0 + (S_{12} - A_{12}A_{22}^{-1}S_{22})(\pi + S_{22})^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

$$= A_{11} - A_{12}Q_{22}^{-1}Q_{21} + (S_{12} + A_{12}Q_{22}^{-1}A_{22}') \times (\pi + S_{22})^{-1} \times (A_{22}Q_{22}^{-1}Q_{21} - A_{21}),$$

$$\hat{B}_0 = B^{\square}\chi^{-1/2},$$

$$\hat{D}_0 = D^{\square},$$

$$\hat{Q} = Q^{\square} - (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})\left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)^{-1}$$
$$\times B_2\chi^{-1}B'_2 \times \left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)^{-1}$$
$$\times (A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$
$$= Q^{\square} - (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})\left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)$$
$$\times B_2B'_2(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$
$$= Q_{11} - Q_{12}Q_{22}^{-1}Q_{21} + (Q_{12}Q_{22}^{-1}A'_{22} - A'_{21})$$
$$(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})$$

Now, we can compute $\hat{S}$ as follows

$$\hat{S}_0 = \hat{B}_0\hat{B}'_0 - \hat{D}_0\hat{D}'_0$$
$$B_0\chi^{-1}B'_0 + \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}B_2B'_0$$
$$+ B_0B'_2(\pi + S_{22})^{-1}D_2D_0\frac{1}{\gamma^2}$$
$$+ \frac{1}{\gamma^2}D_0D'_2\left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)^{-1}B_2B'_2(\pi$$
$$+ S_{22})^{-1}D_2D'_0\frac{1}{\gamma^2} - \frac{1}{\gamma^2}D_0D_0$$
$$- \frac{1}{\gamma^2}D_0D'_2\left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)^{-1}D_2D'_0\frac{1}{\gamma^2}$$
$$= B_0B'_0 - B_0B'_2(\pi + S_{22})^{-1}B_2B'_0$$
$$+ \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}B_2B'_0$$
$$+ B_0B'(\pi + S_{22})^{-1}D_2D'_0\frac{1}{\gamma^2} - \frac{1}{\gamma^2}D_0D'_0$$
$$- \frac{1}{\gamma^2}D_0D'_2(\pi + S_{22})^{-1}D_2D'_0\frac{1}{\gamma^2}$$
$$= S_{11} + A_{12}Q_{22}^{-1}A_{12} - (S_{12} + A_{12}Q_{22}^{-1}A'_{22})$$
$$\times (\pi + S_{22})^{-1}(S_{21} + A_{22}Q_{22}^{-1}A'_{12})$$

We now give the different steps of the derivation that led from (3 42) to (3 28)

$$u^*_{s\gamma} = \chi^{-1}\Big[-(B'_0 + B'\left(\pi - \frac{1}{\gamma}D_2D'_2\right)^{-1}D_2D'_0\frac{1}{\gamma^2}Z_{s\gamma}$$
$$+ B'_2\left(\pi - \frac{1}{\gamma^2}D_2D'_2\right)^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21})x_s\Big]$$
$$- [-B'_0Z_{s\gamma} + B'_2(\pi + S_{22})^{-1}B_2B'_0Z_{s\gamma} + B_2(\pi + S_{22})^{-1}$$
$$\times \left(-D_2D_0\frac{1}{\gamma^2}\right)Z_{s\gamma} + (A_{22}Q_{22}^{-1}Q_{21} - A_{21})]x_s$$
$$= [-B'_0Z_{s\gamma} + B'_2(\pi + S_{22})^{-1}(S_{21} - S_{22}A_{22}^{-1}A'_{12})Z_{s\gamma}$$
$$+ B'_2(\pi + S_{22})^{-1}(A_{22}Q_{22}^{-1}Q_{21} - A_{21}]x_s$$
$$= B'_1Z_{s\gamma}x_s + B_2(\pi + S_{22})^{-1}((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma}$$
$$- (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))x_s$$

Finally, we give the steps that lead from (3 47) to (3 48) for the fast subsystem

$$u^*_{f\gamma}(t) = -B'_2(t)Z_{f\gamma}(t)(x_2 - x_2)$$
$$= -B'_2Z_{f\gamma}(x_2 + A_{22}^{-1}(A_{21}x_1 + B_2u^*_{s\gamma} + D_2w^*_{s\gamma}))$$
$$= -B'_2Z_{f\gamma}x_2 - B_2Z_{f\gamma}A_{22}^{-1}(A_{21} - B_2B'_1Z_{s\gamma}$$
$$+ B_2B'_2(S_{22} + \pi)^{-1}$$
$$\times ((S_{21} + A_{22}Q_{22}^{-1}A'_{22})Z_{s\gamma} - (A_{21}$$
$$- A_{22}Q_{22}^{-1}Q_{21})) + \frac{1}{\gamma^2}D_2D'_1Z_{s\gamma}$$
$$- \frac{1}{\gamma^2}D_2D'_2(\pi + S_{22})^{-1}((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma}$$
$$- (A_{21} - A_{22}Q_{22}^{-1}Q_{21})))x_1$$
$$= -B'_2Z_{f\gamma}x_2 - B'_2Z_{f\gamma}A_{22}^{-1}(A_{21}$$
$$- S_{21}Z_{s\gamma} + S_{22}(S_{22} + \pi)^{-1}$$
$$\times ((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma} - (A_{21}$$
$$- A_{22}Q_{22}^{-1}Q_{21})))x_1$$
$$= -B'_2Z_{f\gamma}x_2 - B'_2Z_{f\gamma}A_{22}^{-1}(A_{21} - S_{21}Z_{s\gamma}$$
$$- \pi(S_{22} + \pi)^{-1}$$
$$\times ((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma} - (A_{21}$$
$$- A_{22}Q_{22}^{-1}Q_{21}))$$
$$+ (S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma}$$
$$- (A_{21} - A_{22}Q_{22}^{-1}Q_{21}))x_1$$
$$= -B'_2Z_{f\gamma}x_2 - B'_2Z_{f\gamma}Q_{22}^{-1}(-A'_{22}(S_{22} + \pi)^{-1}$$
$$\times ((S_{21} + A_{22}Q_{22}^{-1}A'_{12})Z_{s\gamma}$$
$$- (A_{21} - A_{22}Q_{22}^{-1}Q_{21})) + A'_{12}Z_{s\gamma} + Q_{21}))x_1$$

APPENDIX B A USEFUL LEMMA

The following lemma will be used throughout in the derivations of Appendix C

Lemma 3 Let U_1, U_2, V_1 and V_2 be defined as in (4 8)–(4 11) Then the following relationships hold

$$(1)\ A_{22}U_1 + S_{22}V_1 = -S_{21}, \tag{B1}$$
$$(2)\ A_{22}U_2 + S_{22}V_2 = A_{21}, \tag{B2}$$
$$(3)\ Q_{22}U_2 - A'_{22}V_2 = Q_{21}, \tag{B3}$$
$$(4)\ Q_{21}U_1 - A'_{22}V_1 = A_{12}, \tag{B4}$$
$$(5)\ A_{22}U + S_{22}V = A_{21} - S_{21}Z_{s\gamma}, \tag{B5}$$
$$(6)\ Q_{22} - A'_{22}V = Q_{21} + A'_{12}Z_{s\gamma} \tag{B6}$$

Proof

(1)
$$A_{22}U_1 + S_{21}$$
$$= S_{21} + A_{21}Q_{22}^{-1}A_{12} - A_{22}Q_{22}^{-1}A_{22}(S_{22}$$
$$+ A_{22}Q_{22}^{-1}A'_{22})^{-1}(S_{21} + A_{22}Q_{22}^{-1}A'_{12})$$
$$= -S_{22}(S_{22} + A_{22}Q_{22}^{-1}A_{22})^{-1}(S_{21} + A_{22}Q_{22}^{-1}A'_{12})$$
$$= -S_{22}V_1$$

(2)
$$A_{22}U_2 - A_{21}$$
$$= -A_{21} + A_{22}Q_{22}^{-1}Q_{21} + A_{22}Q_{22}^{-1}A'_{22}(S_{22}$$
$$+ A_{22}Q_{22}^{-1}A'_{22})^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21})$$
$$= -S_{22}(S_{22} + A_{22}Q_{22}^{-1}A'_{22})^{-1}(A_{21} - A_{22}Q_{22}^{-1}Q_{21})$$
$$= -S_{22}V_2,$$

(3)
$$Q_{22}U_2 - Q_{21}$$
$$= Q_{21} - Q_{21} + A'_{22}(S_{22} + A_{22}Q_{22}^{-1}A'_{22})^{-1}$$
$$\times (A_{21} - A_{22}Q_{22}^{-1}Q_{21})$$
$$= A'_{22}V_2,$$

(4)
$$Q_{21}U_1 - A'_{12}$$
$$= A'_{12} - A'_{12} - A'_{22}(S_{22} + A_{22}Q_{22}^{-1}A_{22})^{-1}$$
$$\times (S_{21} + A_{22}Q_{22}^{-1}A'_{12})$$
$$= A'_{22}V_1$$

(5) and (6) are direct extensions of (1), (2), (3) and (4)

APPENDIX C VERIFICATION OF SOME USEFUL RESULTS

Proof of the fact that $Z_{11} = Z_{s\gamma}$, $Z_{12} = Z'_c$, $Z_{22} = Z_{f\gamma}$ *satisfy* (2 30)–(2 32)

It is immediate to see that (2.32) is satisfied For (2.31),

$$
\begin{aligned}
\text{LHS} &= A_{21}'Z_{f\gamma} + Z_{s\gamma}A_{12} + (V' + U'Z_{f\gamma})A_{22} \\
&\quad + (Q_{12} - Z_{s\gamma}S_{12}Z_{f\gamma} - (V' + U'Z_{f\gamma})S_{22}Z_{f\gamma} \\
&= (A_{21}' - Z_{s\gamma}S_{12} - V'S_{22})Z_{f\gamma} + Z_{s\gamma}A_{12} \\
&\quad + V'A_{22} + Q_{12} + U'Z_{f\gamma}A_{22} - U'Z_{f\gamma}S_{22}Z_{f\gamma} \\
&= U'A_{22}'Z_{f\gamma} + U'Q_{22} + U'Z_{f\gamma}A_{22} - U'Z_{f\gamma}S_{22}Z_{f\gamma} \\
&= U'(A_{22}'Z_{f\gamma} + Z_{f\gamma}A_{22} + Q_{22} - Z_{f\gamma}S_{22}Z_{f\gamma}) \\
&= 0,
\end{aligned}
$$

where in going from the second to the third step Lemma 3 has been used

For (2.30),

$$
\begin{aligned}
\text{LHS} &= A_{11}'Z_{s\gamma} + A_{21}'(Z_{f\gamma}U + V) + Z_{s\gamma}A_{11} \\
&\quad + (V' + U'Z_{f\gamma})A_{21} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad - (V' + U'Z_{f\gamma})S_{21}Z_{s\gamma} - Z_{s\gamma}S_{12}(Z_{f\gamma}U + V) \\
&\quad - (V' + U'Z_{f\gamma})S_{22}(Z_{f\gamma}U + V) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad - (A_{21}' - Z_{s\gamma}S_{12})(Z_{f\gamma}U + V) \\
&\quad + (V' + U'Z_{f\gamma})(A_{21} - S_{21}Z_{s\gamma} - S_{22}Z_{f\gamma}U - S_{22}V) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad + (U'A_{22}' + V'S_{22})Z_{f\gamma}U \\
&\quad + (A_{21}' - Z_{s\gamma}S_{12})V + (V' \\
&\quad + U'Z_{f\gamma})(A_{22} - S_{22}Z_{f\gamma})U \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad + V'(S_{22}Z_{f\gamma} + A_{22} - S_{22}Z_{f\gamma})U \\
&\quad + (A_{21}' - Z_{s\gamma}S_{12})V + U'(A_{22}'Z_{f\gamma} \\
&\quad + (Z_{f\gamma}A_{22} - Z_{f\gamma}S_{22}Z_{f\gamma})U \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad + (V'A_{22} - U'Q_{22})U \\
&\quad + (A_{21}' - Z_{s\gamma}S_{12})V \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} + Q_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} \\
&\quad - (Q_{12} + Z_{s\gamma}A_{12})U \\
&\quad + (A_{21}' - Z_{s\gamma}S_{12})V \\
&= \bar{A}_0'Z_{s\gamma} + Z_{s\gamma}\bar{A}_0 - Z_{s\gamma}S_0Z_{s\gamma} + \bar{Q} \\
&= 0,
\end{aligned}
$$

where we have used Lemma 3 throughout the simplification

Verification of the fact that $\epsilon = 0$, $\bar{Y}_{11} = Z_{s\gamma}$, $\bar{Y}_{21} = Z_{0f\gamma}U + V$, $\bar{Y}_{22} = Z_{0f\gamma}$ *satisfy* (4 40)–(4 42)

The verification of (4 42) is straightforward For (4 41), using Lemma 3

$$
\begin{aligned}
\text{LHS} &= A_{21}'Z_{0f\gamma} + Z_{s\gamma}A_{12} + V'A_{22} + U'Z_{0f\gamma}A_{22} + Q_{12} \\
&\quad - Z_{s\gamma}B_1B_2'Z_{0f\gamma} \\
&\quad - V'B_2B_2'Z_{0f\gamma} + \frac{1}{\gamma^2}Z_{s\gamma}D_1D_2'Z_{0f\gamma} + \frac{1}{\gamma^2}(V' \\
&\quad + U'Z_{0f\gamma})D_2D_2'Z_{0f\gamma} \\
&= A_{21}'Z_{0f\gamma} + (Z_{s\gamma}A_{12} + V'A_{22} + Q_{12}) \\
&\quad - Z_{s\gamma}S_{12}Z_{0f\gamma} - V'S_{22}Z_{0f\gamma} \\
&\quad + U'\left(Z_{0f\gamma}A_{22} + \frac{1}{\gamma^2}Z_{0f\gamma}D_2D_2'Z_{0f\gamma}\right) \\
&= (A_{21}' - Z_{s\gamma}S_{12} - V'S_{22} - U'A_{22}')Z_{0f\gamma} \\
&\quad + (Z_{s\gamma}A_{12} + V'A_{22} + Q_{12} - U'Q_{22}) \\
&= 0
\end{aligned}
$$

For (4 40), on the other hand, using again Lemma 3,

$$
\begin{aligned}
\text{LHS} &= A_{11}'Z_{s\gamma} - Z_{s\gamma}B_1B_1'Z_{s\gamma} - V'B_2B_1'Z_{s\gamma} \\
&\quad + Z_{s\gamma}A_{11} - Z_{s\gamma}B_1B_1'Z_{s\gamma} - Z_{s\gamma}B_1B_2'V \\
&\quad + (A_{21}' - Z_{s\gamma}B_1B_2' - V'B_2B_2' \\
&\quad + \frac{1}{\gamma^2}D_1D_2')(Z_{0f\gamma}U + V) \\
&\quad + (V' + U'Z_{0f\gamma})\Big(A_{21} - B_2B_1'Z_{s\gamma} \\
&\quad - B_2B_2'V + \frac{1}{\gamma^2}D_2D_1''Z_{s\gamma}\Big) \\
&\quad + \frac{1}{\gamma^2}Z_{s\gamma}D_1D_1'Z_{s\gamma} + Q_{11} + Z_{s\gamma}B_1B_1'Z_{s\gamma} \\
&\quad + V'B_2B_1'Z_{s\gamma} + Z_{s\gamma}B_1B_2'V \\
&\quad + V'B_2B_2'V + \frac{1}{\gamma^2}(V' + U'Z_{0f\gamma})D_2D_2'(Z_{0f\gamma}U + V) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} + Q_{11} \\
&\quad + (A_{21}' - Z_{s\gamma}S_{12} - V'S_{22} \\
&\quad + \frac{1}{\gamma^2}U'Z_{0f\gamma}D_2D_2')(Z_{0f\gamma}U + V) \\
&\quad + V'(A_{21} - S_{21}Z_{s\gamma}) \\
&\quad + U'Z_{0f\gamma}(A_{21} - S_{21}Z_{s\gamma} - B_2B_2'V) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} + Q_{11} \\
&\quad + U'A_{22}'(Z_{0f\gamma}U + V) \\
&\quad + \frac{1}{\gamma^2}U'Z_{0f\gamma}D_2D_2'Z_{0f\gamma}U + V'(A_{21} - S_{21}Z_{s\gamma}) \\
&\quad + U'Z_{0f\gamma}(A_{22}U + S_{22}V - S_{22}V) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} + Q_{11} \\
&\quad + U'(A_{22}'Z_{0f\gamma} + Z_{0f\gamma}A_{22} \\
&\quad + \frac{1}{\gamma^2}Z_{0f\gamma}D_2D_2'Z_{0f\gamma})U + U'A_{22}'V + V'(A_{21} - S_{21}Z_{s\gamma}) \\
&= A_{11}'Z_{s\gamma} + Z_{s\gamma}A_{11} - Z_{s\gamma}S_{11}Z_{s\gamma} + Q_{11} \\
&\quad + U'(\quad Q_{21} - A_{12}'Z_{s\gamma}) \\
&\quad + V'(A_{21} - S_{21}Z_{s\gamma}) \\
&= \bar{A}_0'Z_{s\gamma} + Z_{s\gamma}\bar{A}_0 - Z_{s\gamma}S_0Z_{s\gamma} + \bar{Q} \\
&= 0
\end{aligned}
$$

Verification of the nonsingularity of Ξ_{11}

In Theorem 1, in order to be able to use the Implicit Function Theorem, we needed the condition that Ξ_{11} is nonsingular We now verify this result here First, it is easy to see that Ξ_{11} has to be in the form (4 14), where R is given by

$$
\begin{aligned}
R = A_{11} &- (A_{12} - S_{12}Z_{f\gamma})(A_{22} - Z_{22}Z_{f\gamma})^{-1} \\
&\times A_{21} - S_{11}Z_{s\gamma} - S_{12}(Z_{f\gamma}U + V) \\
&+ (A_{12} - S_{12}Z_{f\gamma})(A_{22} - S_{22}Z_{f\gamma})^{-1}S_{21}Z_{s\gamma} \\
&+ (A_{12} - S_{12}Z_{f\gamma})(A_{22} \\
&- S_{22}Z_{f\gamma})^{-1}S_{22}(Z_{f\gamma}U + V)
\end{aligned}
$$

This can further be simplified to (by using Lemma 3 of Appendix B)

$$
\begin{aligned}
R &= A_{11} - S_{11}Z_{s\gamma} - (A_{12} - S_{12}Z_{f\gamma})(A_{22} \\
&\quad - S_{22}Z_{f\gamma})^{-1}(A_{21} - S_{21}Z_{s\gamma} - S_{22}(Z_{f\gamma}U + V)) \\
&\quad - S_{12}(Z_{f\gamma}U + V) \\
&= A_{11} - S_{11}Z_{s\gamma} - (A_{12} - S_{12}Z_{f\gamma})(A_{22} \\
&\quad - S_{22}Z_{f\gamma})^{-1}(A_{22} - A_{22}Z_{f\gamma})U \\
&\quad - S_{12}(Z_{f\gamma}U + V) \\
&= A_{11} - S_{11}Z_{s\gamma} - A_{12}U - S_{12}V \\
&= \bar{A}_0 - S_0Z_{s\gamma}
\end{aligned}
$$

Since $\bar{A}_0 - S_0Z_{s\gamma}$ is Hurwitz, for $\gamma > \gamma_{s\infty}$, the eigenvalues of Ξ_{11} are in the left half plane From this it follows that Ξ_{11} is nonsingular

For (4 29), on the other hand, R is given by

$$
\begin{aligned}
R = {} & A_{11} - B_1 B_1' Z_{s\gamma} - B_1 B_2' Z_c - (A_{12} \\
& - S_{12} Z_{f\gamma})(A_{22} - S_{22} Z_{f\gamma})^{-1} \\
& \times (A_{21} - B_2 B_1' Z_{s\gamma} - B_2 B_2' Z_c) \\
& + \frac{1}{\gamma^2} D_1 D_1' Z_{s\gamma} + \frac{1}{\gamma^2} D_1 D_2' Z_c \\
& - (A_{12} - S_{12} Z_{f\gamma})(A_{22} - S_{22} Z_{f\gamma})^{-1} \frac{1}{\gamma^2} D_2 D_1' Z_{s\gamma} \\
& - (A_{12} - S_{12} Z_{f\gamma})(A_{22} - S_{22} Z_{f\gamma})^{-1} \frac{1}{\gamma^2} D_2 D_2' Z_c,
\end{aligned}
$$

which simplifies to

$$
\begin{aligned}
R = {} & A_{11} - S_{11} Z_{s\gamma} - S_{12}(Z_{f\gamma} U + V) \\
& - (A_{12} - S_{12} Z_{f\gamma})(A_{22} - S_{22} Z_{f\gamma})^{-1} \\
& \times (A_{21} - S_{21} Z_{s\gamma} - S_{22}(Z_{f\gamma} U + V) \\
= {} & A_0 - S_0 Z_{s\gamma},
\end{aligned}
$$

from which it again follows that Ξ_{11} is nonsingular

Now, for (4 44), R can be rewritten as follows

$$
\begin{aligned}
R = {} & A_{11} - B_1 B_1' Z_{s\gamma} - B_1 B_2' V - (A_{12} \\
& + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma})(A_{22} + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma})^{-1} \hat{A}_{21} \\
& + \frac{1}{\gamma^2} D_1 D_1' Z_{s\gamma} + \frac{1}{\gamma^2} D_1 D_2'(Z_{0f\gamma} U + V) \\
& - (A_{12} + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma})(A_{22} \\
& + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma})^{-1} \frac{1}{\gamma^2} D_2 D_1' Z_{s\gamma} \\
& - (A_{12} + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma})(A_{22} \\
& + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma})^{-1} \frac{1}{\gamma^2} D_2 D_2'(Z_{0f\gamma} U + V) \\
= {} & A_{11} - S_{11} Z_{s\gamma} - S_{12} V + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma} U \\
& - (A_{12} + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma}) \\
& \times \left(A_{22} + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma}\right)^{-1} \left(A_{21} - S_{21} Z_{s\gamma}\right. \\
& \left. - S_{22} V + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma} U\right) \\
= {} & A_{11} - S_{11} Z_{s\gamma} - S_{12} V + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma} U \\
& - \left(A_{12} + \frac{1}{\gamma^2} D_1 D_2' Z_{0f\gamma}\right) \\
& \times (A_{22} + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma})^{-1} \left(A_{22} + \frac{1}{\gamma^2} D_2 D_2' Z_{0f\gamma}\right) U \\
= {} & A_{11} - S_{11} Z_{s\gamma} - S_{12} V - A_{12} U \\
= {} & A_{11} - S_{12} V_2 - A_{12} U_2 - (S_{11} \\
& - S_{12} V_1 + A_{12} U_1) Z_{s\gamma} \\
= {} & \tilde{A}_0 - S_0 Z_{s\gamma},
\end{aligned}
$$

from which, again, the nonsingularity of Ξ_{11} follows

APPENDIX D THEOREM L OF KOKOTOVIC AND YACKEL (1972)

Consider the initial value problem

$$\frac{d\xi}{dt} = f(t, \xi, \eta, \epsilon) \quad \xi(t_0) = \xi_0, \tag{D1}$$

$$\epsilon \frac{d\eta}{dt} = g(t, \xi, \eta, \epsilon) \quad \eta(t_0) = \eta_0, \tag{D2}$$

where ϵ is a small positive parameter, and ξ and η are n- and m-dimensional vectors, respectively Formally setting $\epsilon = 0$ in the system (D1)–(D2) yields the degenerate system

$$\frac{d\bar{\xi}}{dt} = f(t, \bar{\xi}, \bar{\eta}, 0), \quad \bar{\xi}(t_0) = \xi_0, \tag{D3}$$

$$0 = g(t, \bar{\xi}, \bar{\eta}, 0) \tag{D4}$$

Since (D4) may have several roots, suppose that a particular root $\bar{\eta} = \phi(t, \bar{\xi})$ is of interest and substitute it in (D3) The n-dimensional system

$$\frac{d\bar{\xi}}{dt} = f(t, \bar{\xi}, \phi(t, \bar{\xi}), 0), \quad \bar{\xi}(t_0) = \xi_0, \tag{D5}$$

is a reduced system of (D1)–(D2)

Introducing a new time variable τ, (D2) can be rewritten in the form of a boundary layer system,

$$\frac{d\omega}{d\tau} = g(\alpha, \beta, \omega, 0), \quad \omega(t_0) = \eta_0, \tag{D6}$$

where $\alpha = t_0$ and $\beta = \xi(t_0)$ are fixed parameters In the space of variables ξ, η, ϵ, t we define a region $\mathfrak{R}$: $\|\xi - \bar{\xi}(t)\| < r$, $\|\eta - \bar{\eta}(t)\| < r$, $0 \le \epsilon \le \epsilon^0$, $t_0 < t \le t_f$ where $r > |\eta_0 - \bar{\eta}(t_0)| > 0$

Theorem 3 Let the following conditions be satisfied

L1 f, $\partial f/\partial\xi$, $\partial f/\partial\eta$, g, $\partial g/\partial\xi$, $\partial g/\partial\eta$ are of class C^0 in $(\xi, \eta, t, \epsilon) \in \mathfrak{R}$

L2 The solution $\omega(\tau)$ of (D6) exists on $\tau \in [0, \infty)$, is unique, and is asymptotically stable with respect to the root $\phi(t_0, \xi_0)$ of (D4)

L3 The solution $\bar{\xi}(t)$ of the reduced system (D5) exists and is unique on $t \in [t_0, t_f]$

L4 The real parts of the eigenvalues of the Jacobian matrix

$$\frac{\partial g}{\partial \eta}(t, \bar{\xi}, \bar{\eta}, 0)$$

are negative on $[t_0, t_f]$, for $\bar{\eta} = \phi(t, \bar{\xi})$

Then for sufficiently small ϵ, the system (D1)–(D2) has a unique solution $\xi(t, \epsilon)$, $\eta(t, \epsilon)$ on $t \in [t_0, t_f]$ satisfying the initial conditions $\xi(t_0, \epsilon) = \xi_0$, $\eta(t_0, \epsilon) = \eta_0$ Furthermore,

$$\lim_{\epsilon \to 0} \xi(t, \epsilon) = \bar{\xi}(t) \quad \text{on } [t_0, t_f] \tag{D7}$$

$$\lim_{\epsilon \to 0} \eta(t, \epsilon) = \bar{\eta}(t) \quad \text{on } (t_0, t_f], \tag{D8}$$

where the limit (D7) is uniform in t on $[t_0, t_f]$ and the limit (D8) is uniform in t on any interval $(t_1, t_f]$, where $t_0 < t_1 < t_f$

Automatica, Vol 29, No 2, pp 425–429, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Adaptive Cross-direction Control of Paper Basis Weight*

ALENA HALOUSKOVÁ,† MIROSLAV KÁRNÝ† and IVAN NAGY†

Key Words—Adaptive control; computer control; process parameter estimation; control system design; paper industry.

Abstract—The initial setting and maintaining of the desired paper basis weight (in machine direction and cross direction) is to be ensured adaptively for several given paper grades

Technical equipment includes: usual traversing gauge at the output of the paper machine, special traversing robot for shaping the headbox lip (sequential setting of screws), control computer.

Adaptive multivariate LQG control with recursive identification generalized to distributed parameter system has been designed with the following features: integral (convolution type) model of the process respecting the continuous nature of all signals and kernels involved (by means of spline approximation), repetitive control synthesis (via Riccati equation with periodical solutions).

1. *Introduction*

RELIABLE HOMOGENEITY of production in most chemical technologies cannot be ensured without the advanced means of automatic control. It holds also in the paper industry, where the demands on quality distribution over the whole sheet are increasing.

Paper machine performance is evaluated according to three mutually dependent variables prescribed for a given paper grade: basis weight, moisture, and caliper content. Formerly, only the temporal mean properties of the paper sheet (in machine direction—MD) have been controlled automatically; nowadays the desired spatial distribution along the sheet cross-section (in cross direction—CD) is pursued, too. This contribution deals with the CD control of paper basis weight; similar problems are met also in other sheet-making technologies.

The authors belong to a group, where the Linear Quadratic Gaussian (LQG) approach to adaptive control has been deeply elaborated (Kárný *et al.*, 1985) together with the Bayesian approach to identification (Peterka, 1981). The underlying problem formulation was found to fit the requirements of the paper CD control, too.

The contribution begins with a survey of the technology (Section 2). Attention is paid to the discussion of the difficulties met when solving the CD control problem for the basis weight (Section 3). After a short summary of the present state (Section 4), available reserves for improving the quality and economy are suggested (Section 5). The proposed approach is reviewed, stressing the practical aspects more than the theoretical ones, the reasons more than the tools (Section 6). The control system has been broadly tested by simulation; illustrative examples are given in Section 7 The outline of the resulting algorithm follows (Section 8) The results are summarized in the Conclusions (Section 9).

2 *Technological background*

A very simplified scheme of a paper machine is sketched in Fig. 1 to give a non-specialist a rough idea of the design problems.

The raw material, a watery pulp, is pressed out from the headbox through a narrow slice lip onto moving wire section; water flows through the net where paper web is created and then dried. At the output of the machine, paper properties are measured by traversing gauges. The mean machine direction (MD) basis weight is given mainly by the thick stock flow, but small deviations can be controlled by machine velocity. The only way of influencing the cross direction (CD) basis weight is to shape the headbox lip.

3. *Main difficulties*

The main difficulties and expenses are caused by measuring instruments and actuators (continuous signals discretely measured, control actions severely limited, special expensive gauges and devices needed), others are part of technology (multivariate couplings, transportation lag) and possibly of problem formulation (MD and CD control inseparability).

The output evaluation is based on measured data provided by a gauge traversing across the produced paper sheet. The sheet moves at the velocity of the order $10\,\mathrm{m\,sec^{-1}}$ and the measurement principle does not allow the gauge to move more quickly than about $10^{-1}\,\mathrm{m\,sec^{-1}}$; these movements result in a zig-zag measured path sketched in Fig. 2, where the angle α may be about $1°$ only. Moreover, the gauge stops at one side of the sheet (for some time intervals at unpredictable time instants) for recalibrating itself. From these extremely sparse data, the whole profile is to be reliably reconstructed. This is possible only due to relative stationarity of the profile, corrupted, however, by the measurement noise

At the actuator side, the shaping of the slice lip is necessary, which is realized by a set of screws in distance about 0.15 m. For technological reasons, the absolute values of (vertical) screw-position changes and relative adjacent-screw positions are severely limited. The precise screw setting is very demanding and consequently expensive in itself. Position change of one screw afflicts broad area of the web over several adjacent screws (the influence distribution is typical for a given paper grade). This coupling makes the problem essentially multivariate.

The inevitable transportation lag from the headbox (input) to the gauge (output) is substantial.

* Received 24 July 1989; revised 1 October 1990; revised 1 March 1991; received in final form 30 June 1991. The original version of this paper was presented at the 11th IFAC World Congress on Automatic Control of the Service of Mankind which was held in Tallinn, Estonia during August, 1990. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor D. W. Clarke under the directions of Editor P. C. Parks.

† Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Pod vodárenskou věží 4, 182 08 Prague 8, Czechoslovakia.

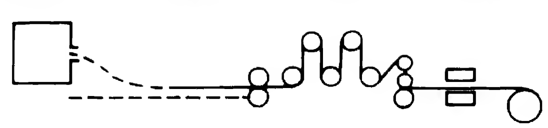


FIG. 1. Scheme of paper machine.

4. *Usual solution and its disadvantages*

The introducing of CD control is explosive in developed countries since 1980, but after the first publications (Karlsson and Hagglund, 1983; Wilhelm and Fjeld, 1983; Chen *et al.*, 1986) rare references appear (e.g. Elliot (1987); Natarajan *et al.* (1988); Brewster (1989)). The information published by the suppliers often does not allow to understand the underlying design principles. In our knowledge, the time–space distribution of the paper basis weight $Y(t, x)$ is usually described (for control purposes) by a variant of the simple model

$$\mathbf{Y}(t) - \mathbf{Y}_0 = \mathbf{Y}_S - \mathbf{Y}_0 + \mathbf{B}_n \mathbf{U}(t - d), \tag{1}$$

where t is discrete time, d is time delay, $'$ denotes transposition, $\mathbf{Y}'(t) = [\mathbf{Y}(t, x_{v1}), \ldots, \mathbf{Y}(t, x_{vv})]$ is a vector of basis-weight values at the points $x \in \{x_{v1}, \ldots, x_{vv}\}$ (usually the CD screw positions), $\mathbf{Y}_0$, $\mathbf{Y}_S$ are desired and starting values of $\mathbf{Y}(t)$, respectively, $\mathbf{U}(t - d)$ is μ-vector of screw settings, $\mathbf{B}_n$ is symmetrical band matrix (the band width is given by the number of screw sections influenced by one screw).

The control task is to remove the deviation caused by an inappropriate initial slice-lip setting. Other disturbances are supposed to be slow, so that the nonzero control action is rarely needed after initial tuning. Machine properties are assumed to be time invariant; the matrix $\mathbf{B}_n$ can be estimated beforehand.

If the technological arrangement allows to change all the screw positions at the same time, it suffices to estimate $\mathbf{Y}_S$ and to set $\mathbf{U}(t) = \mathbf{B}_n^+(\mathbf{Y}_0 - \mathbf{Y}_S)$ for achieving satisfactorily small deviation from the desired output ($^+$ denotes a generalized inversion). The profile is estimated by averaging and filtering the output measured inside the appropriate sheet section.

This solution has several disadvantages:

- the profile estimation adds substantial dynamics to control loops,
- the profile model is very simplified,
- the matrix $\mathbf{B}_n$ has large dimensions and it is difficult to invert it robustly,
- the simultaneous resetting of all screws is expensive.

5. *Available reserves for the economical and qualitative improvements*

The cross-profile basis weight distribution becomes economically important and the reconstruction of old paper machines is an actual task. As measuring instruments and computational means are usually available (and used in the MD control) for large machines, the necessary new hardware is a headbox with slice screw actuators to be controlled directly by a computer program. The simultaneous setting of all screws is expensive; a successive setting using a special robot might be preferred. More complex software is able to make up for this hardware simplification. Improvements of measuring devices and actuators, where possible, seem to be effective. Stationary profile measurement (Francis and Kleinsmith, 1988) and recommendations for slice lip shaping (Kiessling, 1988) have been published recently.

The common definition of the CD profile as a deviation from the MD value makes the MD and CD control loops inevitably interrelated. Changed definition can better reflect control design requirements (see next section).

A substantial quality reserve is still felt in more sophisticated algorithms taking into account technology properties. Profile stationarity is used, but the very profile smoothness has not yet been exploited.

As the control computer is at disposal for the initial tuning, it can be used for adaptive control with recursive identification as well. Nevertheless, the adaptive control still seems to be rare (Ma and Williams, 1988).

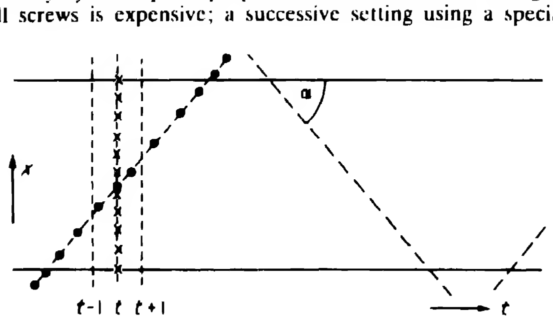


FIG. 2. Measurement path (result of sheet and traversing gauge movement).

6. *The proposed solution*

Problem:

- ensuring possible adherence to the prescribed basis weight profile,
- adding the CD control to the existing MD control.

Hardware available:

- measuring device—traversing radiation gauge,
- actuator—traversing robot (sequential resetting of screws).

6.1. *Control synthesis.* We have found useful to define the profile as a curve shape only (as it is intuitively natural) and to express it in terms of derivatives (cf. the model (1)).

$$\mathbf{s}'(t) = [s_1, \ldots, s_\sigma] = \left[\frac{\partial Y(t, x_{s1})}{\partial x_{s1}}, \ldots, \frac{\partial Y(t, x_{s\sigma})}{\partial x_{s\sigma}}\right], \tag{2}$$

at some fixed CD positions $\{x_{si}\}_{i=1}^{\sigma}$ (usually, derivatives in the middle of screw distance with $\sigma = \mu - 1$). Then the existing MD control can be preserved (as required) while either automatic or manual CD setting is chosen (according to permanent cross profile evaluation and displaying).

The difficulty to synchronize the gauge and actuator movements requires separation of identification and control synthesis, which is well possible in adaptive control design (certainty equivalence principle).

As only single screw-position change is to be realized at one time moment and all future changes must be simultaneously taken into account due to coupling of effects, the long-horizon dynamic programming is the right tool for control synthesis. The optimization criterion is, therefore,

$$\mathcal{K}(t) = \frac{1}{M}\mathcal{E}_{t-1}\left[\sum_{\tau=t}^{t+M-1} l(\tau)\right], \tag{3}$$

where $\mathcal{E}_{t-1}[\cdot]$ is expectation (conditioned on the observed data up to $t - 1$), $l(\tau)$ is partial loss, M is horizon length.

The partial loss $l(t)$ includes the control error; penalization of control actions enables the user to limit them. The quadratic form is advantageous as it emphasizes larger deviations:

$$l(t) = (\mathbf{s}(t) - \mathbf{s}_0)'(\mathbf{s}(t) - \mathbf{s}_0) + w(t)\,\Delta u^2(t), \tag{4}$$

where $\mathbf{s}(t) - \mathbf{s}_0$ is control error (shape error), (σ-vector of the differences between the actual $\mathbf{s}(t)$ and the required profile $\mathbf{s}_0$),

$\Delta u(t)$ is the control action designed (single screw-position change) (see (1))

$$\mathbf{U}'(t) = \mathbf{U}'(t-1) + [\underbrace{0, \ldots, 0, \Delta u(t)}_{i(t)}, 0, \ldots, 0]$$

$$= \sum_{\tau=1}^{t} [\underbrace{0, \ldots, \Delta u(\tau)}_{i(\tau)}, 0, \ldots, 0], \tag{5}$$

$w(t)$ is a time-varying penalization; its restriction-dependent

changes (see Böhm and Kárný, 1982) ensure that slice would not be damaged by excessive input values

As discussed by Kárny *et al* (1988), the profile shape can be described by the state space equation

$$\mathbf{s}(t) = \mathbf{s}(t-1) + \mathbf{f}'^{i(t)}\,\Delta u(t) + \mathbf{e}(t) \qquad (6)$$

where $\mathbf{e}(t)$ is process noise,

$$\mathbf{f}'^{i(t)} = [0, \ldots, 0, \underset{i(t)}{f_1, \ldots, f_m}, f_{m+1}, \ldots, f_{2m}, 0, \ldots, 0]$$

The vector $\mathbf{f}' = [f_1, \ldots, f_{2m}]$ is a skew-symmetrical $k - 2m$ vector ($k \le \sigma$ characterizing the screw influence)

The upper index $i(t)$ points to the screw to be positioned It is periodical and makes the vector $\mathbf{f}'^{i(t)}$ periodical as well (by shifting $\mathbf{f}$ inside $\mathbf{f}'^{i(t)}$)

The equation (6) says that the shape vector $\mathbf{s}(t)$ is a state vector (unmeasurable), the new shape equals the old shape plus an increment following the change of the $i(t)$th screw position $\Delta u(t)$ The increment encompasses more than one section—it is the derivative of the screw response Dynamic programming for this simple state equation (6) and the criterion (3) results in Riccati equation with periodical solutions (enforced by periodicity of $\mathbf{f}'^{i(t)}$) of the form

$$\Delta u(t) = -\mathbf{L}'^{i(t)}\,\mathbf{s}(t-1) \qquad (7)$$

$$\mathbf{L}'^{i(t)} = \frac{\mathbf{S}(t)\mathbf{f}'^{i(t)}}{\mathbf{f}'^{i(t)}\mathbf{S}(t)\mathbf{f}'^{i(t)} + w(t)} \qquad (8)$$

where the matrix $\mathbf{S}(t)$ is the result of backwards recursion

$$\mathbf{S}(\tau) = \mathbf{S}(\tau+1)\left(\mathbf{I} - \frac{\mathbf{f}'^{i(\tau)}\mathbf{f}'^{i(\tau)}\mathbf{S}(\tau+1)}{\mathbf{f}'^{i(\tau)}\mathbf{S}(\tau+1)\mathbf{f}'^{i(\tau)} + w(\tau)}\right) + \mathbf{I} \qquad (9)$$

for $\tau = t + M - 1, \ldots, t$ with the initial condition $\mathbf{S}(t+M) = \mathbf{I}$

6 2 *Identification* The prerequisite for the control approach given is a valid shape estimate $\mathbf{s}(t)$ acquired by identification The data at disposal for this purpose are

- the slice-lip position $\mathbf{U}(t-d, x)$ at time t and CD position $x \in \{x_1, x_2, \ldots, x_\mu\}$, i e (see (1), (5)) $[\mathbf{U}(t-d, x_1), \ldots, U(t-d, x_\mu)] \equiv \mathbf{U}'(t-d)$ (only the deviations from the initial adjustment are available thus we can set $\mathbf{U}(t, \cdot) \equiv 0$ for $t \le 0$),
- the machine speed $V(t-d)$
- the noisy "diagonal" basis weight $Y_d(\vartheta, x(\vartheta))$ measured in the CD positions $x(\vartheta)$ at time moments ϑ (sampling rate is higher than control rate, sampling CD positions are often nonequidistant due to the nonuniform gauge speed see full dots in Fig 2)

The items $\{Y_d(\vartheta, x(\vartheta))\ t-1 < \vartheta \le t\}$ are filtered with a special filter (not adding dynamic delay) (Karny *et al*, 1988) and extrapolated into equidistant points say to the nearest CD screw position $x_{\kappa(t)}$

$$\{Y_d(\vartheta, x(\vartheta))\ t-1 < \vartheta \le t\} \to Y(t, x_{\kappa(t)}), \qquad (10)$$

(see crosses in Fig 2) The prefiltered (smoothed) data are then used

The basic description of the process for the identification and profile display is the convolutional integral equation

$$Y(t, x) = \int_0^L K(\tilde{x})U(t-d, x-\tilde{x})\,\mathrm{d}\tilde{x} + \int_0^t J(\bar{t})V(t-d-\bar{t})\,\mathrm{d}\bar{t} + C(x) + E(t, x), \qquad (11)$$

where $Y(t, x)$ is the basis weight at time t and CD position x, L is width of the paper web, d is transportation lag for the mean speed, $V(t)$ is machine speed at time t, it has to be incorporated into the model because it is used as input variable for the MD basis weight control, $U(t, x)$ is the vertical deviation of slice at time t and CD position x from the initial form, $C(x)$ is a smooth absolute term reflecting slice profile for $U(\cdot, \cdot) \equiv 0$ and $V(\cdot) \equiv V_0 \equiv$ mean machine speed, $K(\cdot)$ is smooth time-invariant even kernel, characteristic for a paper grade, $J(\cdot)$ is smooth time-invariant kernel falling sharply to zero, $E(t, x)$ is stochastic zero-mean term comprising uncertainties

By this model the continuous character of the process in time and space is stressed The references as well as practical experience seem to agree, that the implicit assumptions are fulfilled the system is (approximately) linear static, the edge effects are not pervasive (edges cut off) The assumptions about the disturbance $E(t, x)$ are discussed by Kárný *et al* (1988)

In the evaluations, the continuous functions of a variable h are approximated by properly chosen splines according to formula

$$G(h) = \sum^{n} g_i q_{g\,i}(h) = \mathbf{g}'\mathbf{q}_g(h), \qquad (12)$$

where we have introduced the vector $\mathbf{g}' = (g_1, g_2, \ldots, g_{n_g})$ of the function values, $g_i = G(h_i)$, $i = 1, 2, \ldots, n_g$ at the node points $h_{min} = h_1 < h_2 < \ldots < h_{n_g} = h_{max}$, the vectors of spline bases $\mathbf{q}_g(h)$

First order splines proved to be sufficient Not only the signals, but also integral kernels are approximated (sometimes possibly with nonequidistant spline nodes) In this way the distributed parameter system is approximated by the lumped parameter system

Using the approximation (12) where h stands for x or t, G for C, K, J, U, V, respectively, and $\mathbf{g}$ stands for newly introduced vectors of function values $\mathbf{c}$, $\mathbf{k}$, $\mathbf{j}$, $\mathbf{u}$, $\mathbf{v}$ the equation (11) can be rearranged into the form

$$Y(t, x) = E(t, x) + \mathbf{P}'\mathbf{z}(t, x) \quad x \in [0, L] \qquad (13)$$

The parameter vector $\mathbf{P} = [\mathbf{c}', \mathbf{k}', \mathbf{j}']$ includes the initial profile and parameters of both convolution kernels (discretized according to spline nodes) The kernel parameters $\mathbf{k}$ determine the vector $\mathbf{f}$ (see (6)) used in control synthesis

The regression vector $\mathbf{z}(t, x)$ is a known linear function of $U(t-d, x)$, $V(t-d)$, the weights depend of spline choice only and can be precomputed

Knowing at time t the values $Y(t, x_{\kappa(t)})$, $U(t-d, x_{\kappa(t)})$ (see (10)) the regressor $z(t, x_{\kappa(t)})$ is known, equation (13) can be used for parameter estimation The regression model can be effectively identified by the existing procedures (Kárný *et al*, 1985) including sophisticated forgetting with antibursting influence (Kulhavy 1987) and initialization respecting the prior knowledge (Kárný, 1984)

If the parameters are identified, the basis weight and its derivative, i e the profile shape, can be computed at any CD position (possibly except of spline nodes) By sampling the profile shape at σ CD positions (see (2)) at which derivatives exist, and defining admissible input changes (5), we arrive at the form (6)

7 *System testing*

An extensive system for debugging and testing was built Every part of the algorithm was thoroughly tested without and with noise (much care was devoted to realistic noise simulation)

For the overall system performance the identification is crucial because of sparse data The simultaneous identification of all parameters (13) proved to be reliable but the resolution of mutual influences of $\dim(\mathbf{P}) > \sigma$ parameters requires (much) more than $\dim(\mathbf{P})$, identification periods It was found useful to identify the initial profile first, without input changes During this phase the control algorithm is initialized, too (by iterating the Riccati equation)

The illustrative example shows the performance of a simulated machine modelled according to equation (13) under the following conditions

- number of slice lip screws $\mu = \nu = 15$
- initial absolute profile of the slice lip

$$\mathbf{U} = [0, 0, 0, 0\,1, 0\,2, 0\,3, 0\,2, 0\,1, 0, -0\,1, -0\,2, -0\,3, -0\,2, -0\,1, 0]$$

- sampled screw influence kernel $\mathbf{K}(x) = [0, 0\,7, 1, 0\,7, 0]$

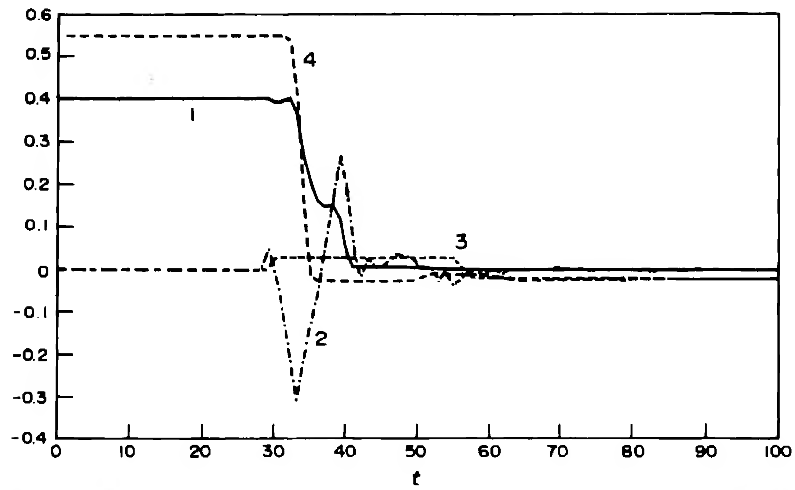


FIG. 3. Closed loop behaviour–time evolutions of: (1) quadratic norm of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$, (2) control action Δu, (3) a typical element of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$, (4) the worst element of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$.

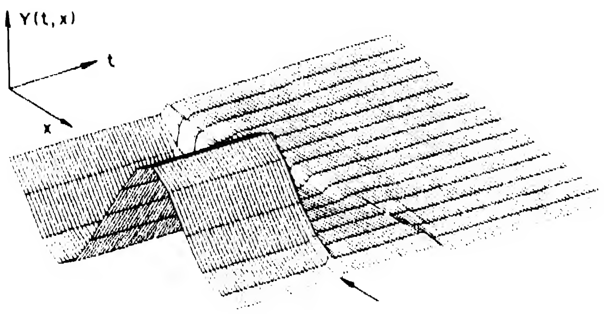


FIG. 4. Time–space distribution of basis weight (known parameters). The arrow (→) marks the start of control.

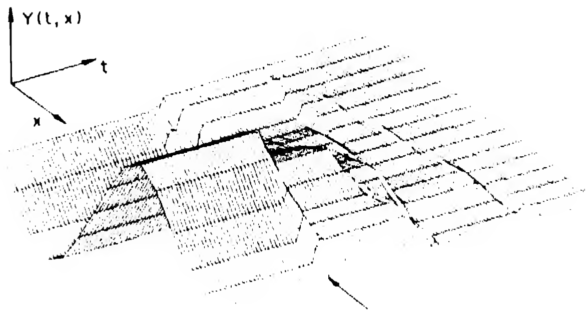


FIG. 6. Time–space distribution of basis weight (unknown parameters). The (→) marks the start of control.

- equidistant spatial spline nodes (placed at the CD screw positions)
- constant velocity

(The noise-less case is illustrated—experiments with noise give satisfactory results, too, but the graphs do not display the control influence clearly enough.)

Figures 3 and 4 show the ideal performance of the system for known parameters and control horizon $M = 30$ longer than one period of the control robot movement. Figures 5 and 6 illustrate how the same system works in case of unknown headbox parameters (velocity fixed). The initial profile is supposed to be completely unknown, the screw-kernel estimate models weak input influence. The resulting strong control inputs in the initial phase are automatically severely limited until the screw kernel is well identified.

8. *The final control algorithm*

The adaptive controller was described by source program parts with user-oriented commentaries. For all parameters appropriate default values were given.

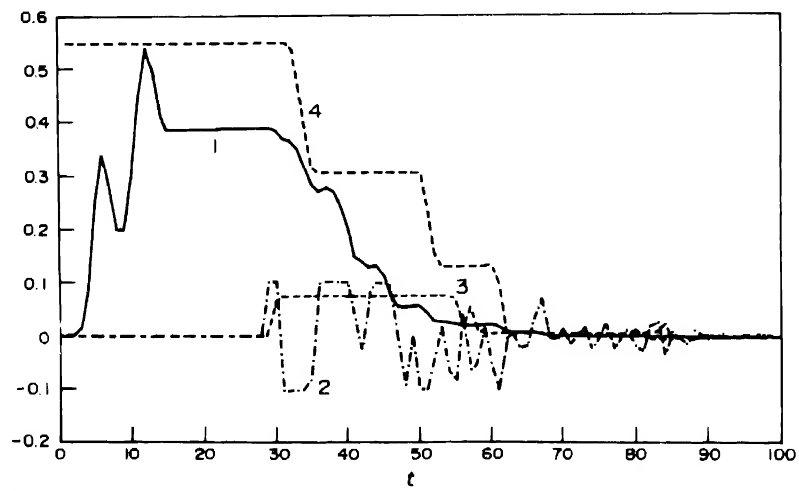


FIG. 5. Closed loop behaviour-time evolutions of: (1) quadratic norm of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$, (2) control action Δu, (3) a typical element of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$, (4) the worst element of the basis weight error $\mathbf{Y}(t) - \mathbf{Y}_0$.

Identification algorithm
Off-line phase:

- spline bases choice (number and placing of nodes; Bayesian structure determination (Kárný and Kulhavý, 1988) might be helpful);
- precomputation of weights for data vector forming (Kárný *et al.*, 1988);
- initialization of recursive least squares (using data from past runs, possibly incorporating prior information (Kárný, 1984)),
- choice of forgetting factor (Kulhavý, 1987).

On-line phase:

- raw data acquisition (without usual filtering which brings additional dynamics);
- filtering (see (10), smoothing and transforming the nonequidistant measurements into equidistant representatives);
- identification of regression model (by recursive least squares with restricted forgetting (Kulhavý, 1987));
- state computation (profile shape in preselected points computed from equation (11) for the current parameter estimates);
- estimation of CD basis weight distribution (display for the staff, again from (11)).

Control algorithm
Off-line phase:

- choice of horizon length (the longer the better, minimally the width of the screw response);
- lower bound of penalization $w(t)$ (3);
- control action limits (given by the user).

On-line phase:

- evaluation of allowed change of actuator action;
- actualization of state dependent restrictions (admissible slice-lip deformation);
- iteration of Ricatti equation (9) (with new parameters);
- evaluation of control action (7);
- hard control signal restriction (if necessary), evaluation of the future penalty w (Böhm and Kárný, 1982).

9. *Conclusions*

Control system for cross-direction paper basis weight control has been designed for sequential resetting of slice lip screws and tested by simulation. Described key results which can be used also in other applications are:

- spline-approximation of convolutional models (Kárný *et al.*, 1990),
- transformation of multi-input–multi-output system into single-input–multi-output with model periodicity.

Special features related to the paper-making implementation are:

- the problem can be formulated as LQG adaptive control problem,
- adaptation to slow process changes is possible,
- the periodical design requires the long-horizon dynamic programming,
- the lack of measured data for identification can be compensated by using prior information (continuity in space and time),
- the experimental results show that the horizon length is critical: it must include at least the width of one screw influential zone,
- the automatically varying penalization (Böhm and Kárný, 1982) proved to be helpful when starting the algorithm under uncertainty.

References

Böhm, J. and M. Kárný (1982). Self-tuning regulators with restricted inputs. *Kybernetika*, **18**, 529–544.

Brewster, D. B. (1989). Profile control by distributed control systems: dream or reality? Tappi J., **72**, 275–281.

Elliot, R (1987). Selecting CMD (cross-machine direction) controls to get exceptional results. *ARPC Congr. Annual 20th*, Sao Paulo, 16–20 November, pp. 327–333.

Francis, K. and Ch. Kleinsmith (1988). The stationary bone-dry weight sensor. *Tappi Proc. Eng. Conf.*, 88, Chicago IL, 19–22 September, **2**, 561–567.

Chen, S. C., R. M. Snyder and R. G. Wilhelm (1986). Adaptive profile control for sheetmaking processes. *6th Int. IFAC/IMEKO Conf. on Instrum. and Autom. in the Paper, Rubber, Plastics and Polymerization Industries*, Akron, OH, pp. 77–83.

Karlsson, H and L. Hagglund (1983). Optimal cross-direction basis weight and moisture profile control on paper machines. *3rd Int. Pulp and Paper Process Control, Symposium*, Vancouver, B.C., Canada, pp. 139–145.

Kárný, M. (1984). Quantification of prior knowledge about global characteristics of linear normal model. *Kybernetika*, **20**, 376–385.

Kárný, M., A. Halousková, J. Böhm, R. Kulhavý and P. Nedoma (1985). Design of linear quadratic adaptive control: theory and algorithms for practice. *Kybernetika*, **21** (supplement to numbers 3–6).

Kárný, M., A. Halousková and I. Nagy (1988). Modelling, identification and adaptive control of cross-direction basis weight of paper sheets. *Int. Conf. Control* 88, Oxford, pp. 159–164.

Kárný, M. and R. Kulhavý (1988). Structure determination of regression-type models for adaptive prediction and control. In J. C. Spall (Ed.), *Bayesian Analysis of Time Series and Dynamic Models*. Marcell Dekker, New York.

Kárný, M., I. Nagy, J. Böhm and A. Halousková (1990). Design of spline-based self-tuners. *Kybernetika*, **26**, 17–30.

Kiessling, H. (1988). Anforderungen an den Stoffauflauf Flächengewichts-Querprofilregelung. *Wochenbl. für Papierfabrikation*, **116**, 967–972.

Kulhavý, R. (1987). Restricted exponential forgetting in real-time identification. *Automatica*, **23**, 598–600.

Ma, M. S. and D. C. Williams (1988). A simplified adaptive mode predictive control. *Tappi J.*, **71**, 189–193.

Natarajan, K., A. G. Dumont and M. S. Davies (1988). An algorithm for estimating cross and machine direction moisture profiles for paper machines. *IFAC Symposium on Identification and Process Parameter Estimation*, Beijing, China.

Peterka, V. (1981). Bayesian system identification. In P. Eykhoff (Ed.), *Trends and Progress in System Identification*, pp. 239–304, Pergamon, Oxford.

Wilhelm, R. G. and M. Fjeld (1983). Control algorithms for cross directional control. *5th Int. IFAC/IMEKO Conf. on Instrum. and Autom. in the Paper, Rubber, Plastics and Polymerization Industries*, Antwerp, Belgium, pp. 163–174.

Automatica, Vol. 29, No. 2, pp. 431–438, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Extended Discrete-time LTR Synthesis of Delayed Control Systems*†

JENNY H. SHEN‡ and ASOK RAY‡

Key Words—Discrete systems; optimal control; robustness; delays.

Abstract—This paper addresses compensation of dealys within a multi-input–multi-output discrete-time feedback loop for application to real-time distributed control systems. The delay compensation algorithm, formulated, in this paper, is an extension of the standard loop transfer recovery (LTR) procedure from one-step prediction to the general case of p-step prediction ($p \geq 1$). It is shown that the steady-state minimum-variance filter gain is the H_2-minimization solution of the relative error between the target sensitivity matrix and the actual sensitivity matrix for p-step prediction ($p \geq 1$). This concept forms the basis for synthesis of robust p-step delay compensators ($p > 1$). The proposed control synthesis procedure for delay compensation is demonstrated via simulation of the flight control system of an advanced aircraft.

1. *Introduction*

IN MANY REAL-TIME distributed control systems such as advanced aircraft, spacecraft, and autonomous manufacturing plants, the sensor and control signals within a feedback loop may be delayed or interrupted. An example is the randomly varying delays induced by multiplexed data communication networks in Integrated Communication and Control Systems (ICCS) (Halevi and Ray, 1988; Ray and Halevi, 1988; Liou and Ray, 1990) where delays may be distributed between the sensor and the controller and between the controller and the actuator as illustrated in Fig. 1. Another example is the occurrence of delays in the control law execution due to priority interruption at the control computer (Belle Isle, 1975). In general, the presence of randomly varying distributed delays within a multi-input–multi-output (MIMO) feedback system makes the task of controller design significantly more difficult than that without delays. To this effect Luck and Ray (1990) proposed a delay compensator to alleviate the detrimental effects of randomly varying distributed delays by using a multi-step predictor. The key idea in this multi-step compensator design is to monitor the data when it is generated and to keep track of the delay associated with it. With this knowledge, the problem of varying distributed delays can be alleviated by having a lumped constant delay of multiple sampling intervals as seen by the controller.

The major assumption in the formulation of the above multi-step compensation algorithm (Luck and Ray, 1990) is that the randomly varying delays are bounded. This assumption is justified in view of the fact that an unbounded delay would render the closed loop open. Using a specified confidence interval, an upper bound can be assigned to each of the randomly varying distributed delays. The number (p) of predicted steps in the compensator is then determined from these bounds. That is, at time k, the predictor estimates the state using the measurements up to the $(k - p)$th instant. Although Luck and Ray (1990) addressed some of the robustness issues of the delay compensator for structured uncertainties, the compensated system used the gain matrices that were originally designed for the non-delayed system. Since the robustness property of linear quadratic optimal regulators (LQR) is not retained when the state feedback is replaced by state estimate feedback (Doyle and Stein, 1979), this problem is likely to become worse with the insertion of a p-step predictor for $p \geq 1$ because of the additional dynamic errors resulting from plant modeling uncertainties and disturbances.

The objective here is to develop a procedure for synthesis of the p-step delay compensator with a trade-off between performance and stability robustness. To achieve this goal we propose to extend the concept of loop transfer recovery (LTR) (Doyle and Stein, 1981; Stein and Athans, 1987) which is a well-established procedure for synthesis of robust controllers. In continuous-time systems, the key step in LTR design is to select an observer gain so that the full-state feedback loop transfer property can be recovered asymptotically. For the discrete-time LTR, Maciejowski (1985) has shown that although the target sensitivity matrix can be completely recovered with *a posteriori* state estimation (i.e. $p = 0$ in the p-step predictor) a predictive state estimator (i.e. $p = 1$) is not capable of full recovery. Along this line Zhang and Freudenberg (1991) have analysed the loop transfer recovery error for predictive state estimation. We have adopted an approach, following the multi-step prediction method of Luck and Ray (1990) to synthesize the control system for delay compensation. This approach minimizes the loop recovery error where the gain of the p-step observer is tuned to a prescribed value. An alternative approach is to incorporate the delay in the plant model and then synthesize the controller and observer gains that must accommodate the effects of loss of robustness margins due to the delay and uncertainties. This approach is also discussed in this paper and it has been shown that the above two approaches yield identical relative sensitivity errors. However, if the plant model is simply augmented to accommodate the induced delays, the state-space realization may not be minimal as pointed out by Kinnaert and Peng (1990).

The paper is organized in five sections including the introduction. Section 2 summarizes the pertinent properties of LTR for one-step prediction as reported in the existing literature. Main results including the structure and properties of the p-step delay compensator ($p \geq 1$) are generated in Section 3. A general description of the p-step compensator is first presented. Then, the loop transfer matrix of the compensator is derived along with the error of the sensitivity matrix relative to that of the target loop. Finally, it is shown

* Received 17 September 1991; revised 24 March 1992, received in final form 15 June 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor T. Başar under the direction of Editor H. Kwakernaak.

† This work reported in this paper was supported in part by Office of Naval Research under Grant No. N00014-90-J-1513.

‡ Mechanical Engineering Department, Pennsylvania State University, University Park, PA 16802, U.S.A.

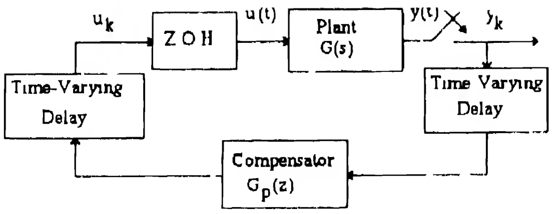


FIG 1 Random distributed delays in the control system

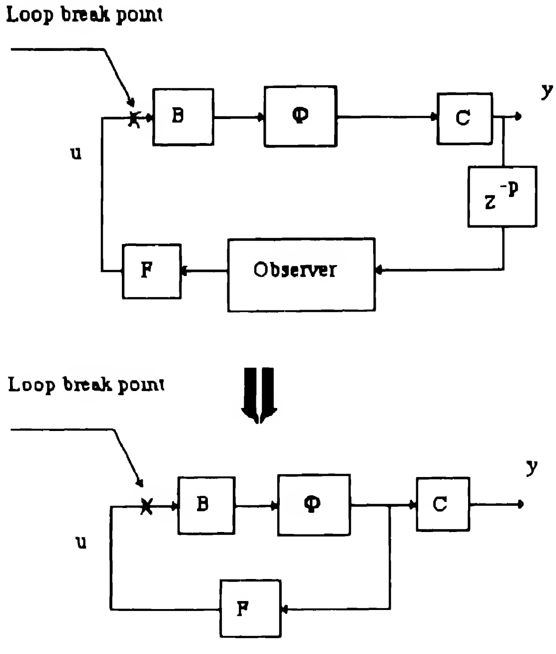


FIG 2 Loop transfer recovery breaking the loop at plant input

that the minimum variance filter gain minimizes the error of sensitivity matrices A procedure to synthesize the observer and controller gains for the p-step compensator is presented in Section 4 The proposed procedure is tested by the simulation of the flight control system of an aircraft which is subjected to a lumped delay of two sampling periods The paper is summarized and concluded in Section 5

2 *Review of the LTR concept for one step prediction*

The concept of loop transfer recovery (LTR) and the existing results for one-step prediction in the discrete time setting are presented in this section The plant under control is represented by a discretized version of a finite dimensional linear time-invariant model in the continuous time setting The discretized model is assumed to be minimum-phase stabilizable and detectable

$$x_{k+1} = Ax_k + Bu_k \qquad y_k = Cx_k \tag{1}$$

Following (1) the plant transfer matrix is given as

$$G(z) = C\Phi(z)B \tag{2}$$

where $\Phi(z) = (zI - A)^{-1}$ is the resolvement matrix The target loop transfer matrix at the plant input is

$$H(z) = F\Phi(z)B \tag{3}$$

and the resulting target sensitivity matrix is

$$S(z) = [I + H(z)]^{-1} \tag{4}$$

The full-state feedback control law for the above plant is

$$u_k = -Fx_k \tag{5}$$

In this paper we have assumed that the uncertainties are unstructured and lumped at the plant input in the form of an input multiplicative term

$$\tilde{G}(z) = G(z)[I + \Delta(z)] \tag{6}$$

with given bound

$$\bar{\sigma}[\Delta(e^{j\omega})] \le l_m(\omega) \quad \forall\omega > 0 \tag{7}$$

Usually, unstructured uncertainties include high frequency dynamics that are not modeled in the plant dynamics The bound $l_m(\omega)$ finally restricts the system design specifications for stability robustness in terms of the closed loop complementary sensitivity matrix (Doyle and Stein 1981) The task of control synthesis via the standard LQG/LTR approach is focused on shaping the loop sensitivity matrices for required performance and stability robustness and can be carried out in two stages as discussed in Doyle and Stein (1979)

For the filter observer (i e $p = 0$) of a stabilizable detectable and minimum phase plant the loop transfer and sensitivity matrices have been shown by Maciejowski (1985) to converge pointwise in frequency to those of the target system as the measurement noise approaches zero However this may not be valid for the one-step predictor (Maciejowski, 1985, Zhang and Freudenberg, 1991, Yen and Horowitz, 1989) Breaking the loop at the plant input, as shown in Fig 2, the one-step delay compensator transfer matrix is

$$G_1(z) = F(zI - A + BF + LC)^{-1}L \tag{8a}$$

An alternative form is

$$G_1(z) = F[I + \Phi(z)(BF + LC)]^{-1}\Phi(z)L \tag{8b}$$

Then the loop transfer matrix for the one step delay compensated system is

$$L_1(z) = G_1(z)G(z) = F[I + \Phi(z)(BF + LC)]^{-1}\Phi(z)LC\Phi(z)B \tag{9a}$$

which can also be expressed, similar to the formula proposed by Zhang and Freudenberg (1991) as

$$L_1(z) = [I + E_1(z)]^{-1}[H(z) - E_1(z)] \tag{9b}$$

where $E_1(z) = F[zI - A + LC]^{-1}B$ is the one step error matrix at the plant input The resulting one step sensitivity matrix can be expressed as a function of the error transfer matrix

$$\begin{aligned} S_1(z) &= [I + L_1(z)]^{-1} \\ &= [I + H(z)]^{-1}[I + E_1(z)] \end{aligned} \tag{10}$$

It is clear from (6) and (9) that $E_1(z)$ is essentially the relative error of the sensitivity matrix $S_1(z)$ of one-step delay compensator relative to the target sensitivity matrix $S(z)$

$$E_1(z) = S(z)^{-1}[S_1(z) - S(z)] \tag{11}$$

It is known (Zhang and Freudenberg, 1991, Yen and Horowitz 1989) that complete loop recovery, i e making $E_1(z) = 0$ for all z cannot be achieved in general by a constant observer gain L However, it is possible to identify an L that minimizes the one-step error transfer matrix $E_1(z)$ in the H_2 sense

3 *The p-step delay compensator*

Figure 1 illustrates a feedback control system where both sensor and control signals are delayed (This might happen in a network based integrated control system of advanced aircraft where sensor, controller and actuator are not collocated) If the sum of the upper bounds of distributed delays are represented by a lumped delay of p sampling intervals at the plant output, the sensory information available at the controller is y_{k-p} at the kth instant The p-step delay compensator (where the plant is completely controllable and observable), proposed by Luck and Ray

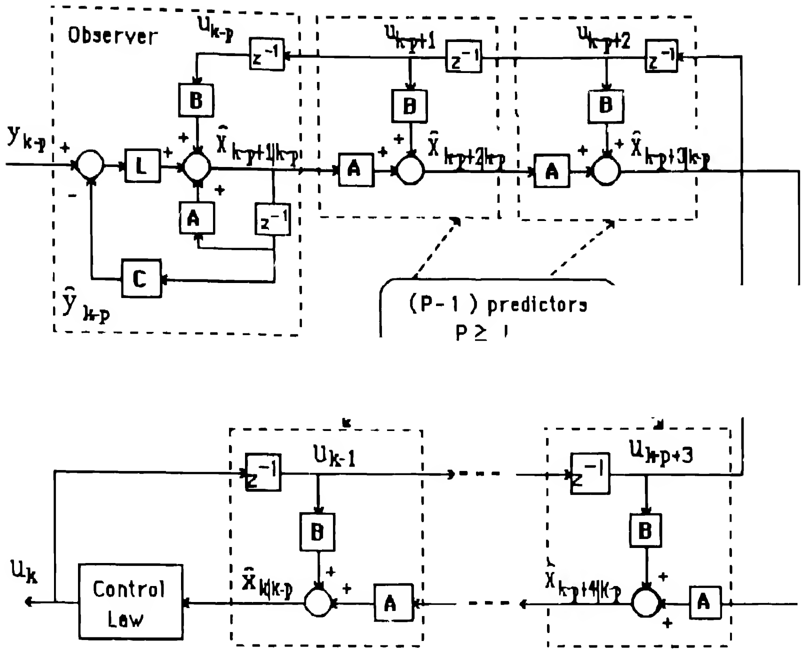


FIG. 3 Structure of p-step delay compensator.

(1990) and illustrated in Fig. 3, has the following structure.

$$u_k = -F\hat{x}_{k|k-p}, \tag{12}$$

where F is the state feeback gain matrix and the state estimate is based on the sensory information up to the $(k-p)$th instant given as:

$$\begin{aligned}\hat{x}_{k|k-p} &= A\hat{x}_{k-1|k-p} + Bu_{k-1}\\ &\vdots\\ \hat{x}_{k-p+2|k-p} &= A\hat{x}_{k-p+1|k-p} + Bu_{k-p+1}\\ \hat{x}_{k-p+1|k-p} &= A\hat{x}_{k-p|k-p-1} + Bu_{k-p} + L(y_{k-p} - C\hat{x}_{k-p|k-p-1}).\end{aligned} \tag{13}$$

The key idea of applying the LTR approach to the above p-step delay compensator is to tune the loop transfer matrix such that the error transfer matrix (i.e. the different between the actual and target sensitivity matrices) is minimized in a certain sense. Derivation of the loop transfer function of the p-step delay compensator is presented below as two propositions.

Proposition 1. The transfer matrix of the p-step delay compensator ($p \geq 1$) from y_k to u_k in the equation set (13) is given as:

$$G_p(z) = F\Omega_p^{-1}(z)\frac{A^{p-1}}{z^{p-1}}\left[zI - A + LC + BF\Omega_p^{-1}(z)\frac{A^{p-1}}{z^{p-1}}\right]^{-1}L, \tag{14}$$

where

$$\Omega_p(z) = I + \left(I - \frac{A^{p-1}}{z^{p-1}}\right)\Phi(z)BF \quad \text{for } p > 1, \quad \text{and } \Omega_1(z) = I. \tag{15}$$

Proof of Proposition 1. The transfer matrix of the p-step ($p > 1$) compensator from y_k to u_k is obtained by substituting (13) into (12):

$$u_k = -FA^{p-1}\hat{x}_{k-p+1|k-p} - F\sum_{i=0}^{p-2} A^iBu_{k-i-1},$$

where the summation on the right hand side reduces to zero for $p = 1$. From the above equation, it follows that

$$u_k = -FA^{p-1}(A-LC)\hat{x}_{k-p|k-p-1} - FA^{p-1}Ly_{k-p} - F\sum_{i=0}^{p-1} A^iBu_{k-i-1}. \tag{16}$$

The Z-transform of (16) is.

$$U(z) = -FA^{p-1}(A-LC)\hat{X}(z)z^{-p} - FA^{p-1}LY(z)z^{-p} - F\sum_{i=0}^{p-1} A^iBU(z)z^{-i-1}, \tag{17}$$

and Z-transform of the last equation in (13) yields:

$$\hat{X}(z) = (zI - A + LC)^{-1}[BU(z) + LY(z)], \tag{18}$$

where $\hat{X}(z) = Z[\hat{x}_{k|k-1}]$, $Y(z) = Z[y_k]$, and $U(z) = Z[u_k]$.

Substituting (18) into (17) yields:

$$\begin{aligned}U(z) = &-FA^{p-1}(A-LC)(zI-A+LC)^{-1}[BU(z)+LY(z)]z^{-p}\\ &-FA^{p-1}LY(z)z^{-p},\\ &-F\sum_{i=0}^{p-2} A^iBU(z)z^{-i-1},\end{aligned}$$

which in turn can be simplified as:

$$\begin{aligned}U(z) = &-F\Big[z^{p-1}I + \sum_{i=0}^{p-2} A^iBz^{p-i-2}F\\ &+A^{p-1}(zI - A + LC)^{-1}BF\Big]^{-1}\\ &\times A^{p-1}(zI - A + LC)^{-1}LY(z).\end{aligned} \tag{19}$$

Using the relationship $\Phi(z) = (zI - A)^{-1}$ and equating like powers of A, equation (15) can be expressed as:

$$\Omega_p(z) = \begin{cases} I + \sum_{i=0}^{p-2} A^iBz^{-i-1}F & p > 1,\\ I & p = 1.\end{cases} \tag{20}$$

The proof is completed by substituting (20) into (19) and exercising a few algebraic operations.

An alternative proof of Proposition 1 can be formulated following (Ishihara, 1988) by expressing the transfer matrix of the p-step compensator from y_k to u_k as:

$$\begin{aligned}G_p(z) = &-\left[\frac{FA^{p-1}}{z^{p-1}}(zI - A + LC)^{-1}B + I + \sum_{i=0}^{p-2}\frac{FA^iB}{z^{i+1}}\right]^{-1}\\ &\times\frac{FA^{p-1}}{z^{p-1}}(zI - A + LC)^{-1}L.\end{aligned} \tag{21}$$

Proposition 2. Let the loop transfer matrix of the p-step

delay compensated system at the plant input be expressed as

$$L_p(z) = G_p(z)G(z),$$

where $G_p(z)$ and $G(z)$ are defined in (14) and (2), respectively Then,

$$L_p(z) = [I + E_p(z)]^{-1}[H(z) - E_p(z)], \tag{22}$$

where

$$E_p(z) = F\Phi(z)B - F\frac{A^{p-1}}{z^{p-1}}[I + \Phi(z)LC]^{-1}\Phi(z)LC\Phi(z)B, \tag{23}$$

is the p-step error transfer matrix, and $H(z)$ is the target loop transfer matrix as defined in (3)

Proof of Proposition 2 By Proposition 1 the open loop transfer matrix of the p-step delay compensator is

$$\begin{aligned}
L_p(z) &= G_p(z)G(z),\\
&= F\Omega_p^{-1}(z)\frac{A^{p-1}}{z^{p-1}}\left[zI - A + LC + BF\Omega_p^{-1}(z)\frac{A^{p-1}}{z^{p-1}}\right]^{-1}\\
&\quad \times LC\Phi(z)B\\
&= F\Omega_p^{-1}(z)\left[I + \frac{A^{p-1}}{z^{p-1}}(I + \Phi(z)LC)^{-1}\Phi(z)\right.\\
&\quad \left.\times BF\Omega_p^{-1}(z)\right]^{-1}\frac{A^{p-1}}{z^{p-1}}\\
&\quad \times (I + \Phi(z)LC)^{-1}\Phi(z)LC\Phi(z)B\\
&= F\left[I + \left(I - \frac{A^{p-1}}{z^{p-1}}\right)\Phi(z)BF + \frac{A^{p-1}}{z^{p-1}}\right.\\
&\quad \left.\times (I + \Phi(z)LC)^{-1}\Phi(z)BF\right]^{-1}\frac{A^{p-1}}{z^{p-1}}\\
&\quad \times (I + \Phi(z)LC)^{-1}\Phi(z)LC\Phi(z)B\\
&= \left[I + F\Phi(z)B - F\frac{A^{p-1}}{z^{p-1}}\right.\\
&\quad \left.\times (I + \Phi(z)LC)^{-1}\Phi(z)LC\Phi(z)B\right]^{-1}\\
&\quad \times F\frac{A^{p-1}}{z^{p-1}}(I + \Phi(z)LC)^{-1}\Phi(z)LC\Phi(z)B
\end{aligned}$$

The proof is completed by substituting (24) and (3) in the above equation

Remark 1 After some algebraic manipulations $E_p(z)$ can be written as

$$E_p(z) = FT_p(z) \tag{24}$$

where

$$T_p(z) = \begin{cases} \dfrac{A^{p-1}}{z^{p-1}}(zI - A + LC)^{-1} + \sum_{i=0}^{p-2} A^iBz^{-i-1} & p \geq 2 \\ (zI - A + LC)^{-1}B & p = 1 \end{cases} \tag{25}$$

This shows that $E_p(z)$ can be separated in terms of the full-state feedback gain F and a function of L and p

Remark 2 For a minimum-phase plant (A, B, C) and with $\det(CB) \neq 0$, as the measurement noise convariance matrix R approaches to zero and consequently $L \to AB(CB)^{-1}$ following Shaked (1985) the error matrix $E_p(z)$ in (23) in Proposition 2 simplifies to

$$E_p(z) = F\left(I - \frac{A^p}{z^p}\right)\Phi(z)B \tag{26}$$

For $p = 1$, the error matrix $E_p(x)$ in (26) is identically equal to $E_1(z)$ in (11)

Remark 3 It follows from the expression of $L_p(z)$ in Proposition 2 that the sensitivity matrix, $S_p(z)$, of the p-step delay compensator is

$$S_p(z) = [I + L_p(z)]^{-1}, \tag{27}$$

and the difference between the sensitivity matrices of the p-step compensated and target systems is

$$S_p(z) - S(z) = [I + H(z)]^{-1}E_p(z), \tag{28}$$

where $S(z)$ is given in (4) This shows that the error transfer matrix $E_p(z)$ is indeed the error of the sensitivity matrix of the p-step delay compensator loop relative to that of the target loop

Remark 4 If the plant model has an inherent delay of p_1 steps and the induced delay in the feedback loop amounts to p_2 steps such that the total delay is $p = p_1 + p_2$ then the resulting error transfer matrix satisfies equation (26) as the measurement noise covariance is tuned to zero This can be easily seen if the plant transfer matrix satisfies the following constraints

$$CA^iB = 0 \quad i = 0, 1, \ldots, p_1 - 2 \text{ and } \det(CA^{p_1-1}B) \neq 0 \tag{29}$$

$$\begin{aligned}
G(z) &= C\Phi(z)B\\
&= C\sum_{i=0}^{\infty}\frac{A^i}{z^{i+1}}B\\
&= C\frac{A^{p_1-1}}{z^{p_1-1}}\Phi(z)B
\end{aligned} \tag{30}$$

According to Shaked (1985) the observer gain L approaches $A^{p_1}B(CA^{p_1-1}B)^{-1}$ as the measurement noise covariance approaches zero Applying this observer gain to the loop transfer matrix

$$L_p(z) = G_p(z)G(z)$$

and by using (14) in Proposition 1 the error matrix becomes identical to that in (26) with a total delay of $p = p_1 + p_2$

This shows that any inherent delay in the plant has the same effect on the error matrix as the induced delay in the feedback loop In other words we can either consider the delays in the feedback loop outside the plant or as a part of the plant model In the second case the original plant state space matrices (A, B, C) with $\det(CB) \neq 0$ need to be augmented with p_1 steps of delay and the new plant state space matrices (A, B, C) need to be formed which must satisfy the following conditions (i) $C(zI - A)^{-1}B = C(zI - A)^{-1}Bz^{-p_1}$ and (ii) complete controllability and observability Therefore we have adopted the first approach of putting the lumped induced delay outside the plant model

Remark 5 Dual results of Proposition 2 obtained by breaking the loop at the plant output instead of plant input yield the loop transfer matrix

$$\begin{aligned}
L_f(z) &= G(z)G_f(z)\\
&= [H(z) - E_p(z)][I + E_p(z)]^{-1},
\end{aligned}$$

where the target loop transfer matrix at the plant output (i e the transfer matrix of the minimum variance filter) is

$$H(z) = C\Phi(z)L$$

and

$$E_p(z) = C\Phi(z)L - C\Phi(z)BF\Phi(z)[I + BF\Phi(z)]^{-1}\frac{A^{p-1}}{z^{p-1}}L$$

The resulting loop sensitivity matrix of the p-step delay compensator at the plant output is

$$S_f(z) = [I + L_f(z)]^{-1} = [I + E_p(z)][I + H(z)]^{-1},$$

and the difference between the sensitivity matrices of the minimum variance filter and p step delay compensator is

$$S_p(z) - S(z) = E_p(z)[I + H(z)]^{-1}$$

H_2-minimization of the p-step error matrix

For the one-step predictor, it has been shown (Zhang and Freudenberg, 1991, Yen and Horowitz, 1989) that the steady-state minimum-variance filter gain with zero measurement noise is obtained by minimizing the H_2 norm of the one-step error matrix $E_1(z)$ Analogous to the case of $p = 1$,

we will show that the same filter gain minimizes the H_2 norm of the p-error matrix $E_p(z)$. This result forms the basis for synthesis of robust p-step delay compensators ($p > 1$) and is presented below as two propositions. Then, the procedure for synthesizing the observer gain matrix is outlined.

For the purpose of tuning the minimum variance gain of the observer, we augment the discrete time, linear, time-invariant plant model in (1) and (2) with (fictitious) plant and measurement noises as

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{31}$$

$$y_k = Cx_k + v_k \tag{32}$$

where $\{w_k\}$ is a zero-mean white sequence with covariance matrix $E\{w_k w_j^T\} = BB^T\delta_{kj}$, and $\{v_k\}$ is a zero-mean white sequence with covariance matrix $E\{v_k v_j^T\} = \rho I\delta_{kj}$, combining the distributed delays within the control loop as a lumped delay of p sampling intervals at the sensor controller interface, the state estimate is redefined as

$$\hat{x}_{k|k-p} = E\{x_k \mid y_{k-p}\} \tag{33}$$

where the estimator is described by the set of equations (13).

Proposition 3. Let the (zero mean) state estimation error be defined as

$$e_{k|k-p} = x_k - \hat{x}_{k|k-p} \tag{34}$$

Then,

$$E\{e_{k|k-p}e^T_{k|k-p}\} = \begin{cases} \displaystyle\sum_{i=0}^{+\infty} A^{p-1}(A-LC)^iBB^T(A-LC)^{iT}A^{(p-1)T} + \sum_{i=0}^{p-2} A^iBB^TA^{iT} & p \ge 2 \\ \displaystyle\sum_{i=0}^{+\infty} A^{p-1}(A-LC)^iBB^T(A-LC)^{iT}A^{(p-1)T} & p = 1 \end{cases} \tag{35}$$

Proof of Proposition 3. From the plant model in (31) and (32) and the filter equations (13) we can express the estimation error $e_{k|k-p}$ in terms of the input sequence $\{w_i\}$ when the system is initially started at $i = -\infty$

$$e_{k|k-p} = \sum_{i=-\infty}^{k-p} A^{p-1}(A-LC)^{k-p-i}w_i + \sum_{i=k-p+1}^{k-1} A^{k-1-i}w_i \tag{36}$$

Hence the cross covariance of e_{k+s} and w_k defined as $R_{ew}(s) = E\{e_{k+s|k+s-p}w_k^T\}$ can be expressed in the following form

$$\begin{aligned} R_{ew}(s) &= \sum_{m=p}^{+\infty} A^{p-1}(A-LC)^{s-p-m}BB^T\delta(m) + \sum_{m=1}^{p-1} A^{s-1-m}BB^T\delta(m) \\ &= \begin{cases} A^{p-1}(A-LC)^{s-p}BB^T & (s \ge p) \\ A^{s-1}BB^T & (s \ge 1, \; p > 1) \\ 0 & (s = 0) \end{cases} \end{aligned} \tag{37}$$

Similarly, the autocovariance of e_k is obtained as

$$R_{ee}(s) = \begin{cases} \displaystyle\sum_{m=0}^{+\infty} R_{ew}(s+m+p)(A-LC)^{mT}A^{(p-1)T} + \sum_{i=0}^{p-2} R_{ew}(s+i+1)A^{iT} & p \ge 2 \\ \displaystyle\sum_{m=0}^{+\infty} R_{ew}(s+m+p)(A-LC)^{mT}A^{(p-1)T} & p = 1 \end{cases} \tag{38}$$

The proof is completed by setting $s = 0$ and then substituting (37) into (38).

Proposition 4. The H_2 norm of the sensitivity error matrix $E_p(z)$, defined in (23) in Proposition 2, is minimized if the observer gain matrix L is identically equal to the standard steady-state minimum-variance filter gain matrix with no measurement noise.

Proof of Proposition 4. From the definition of H_2 norm (Francis, 1987) and $E_p(z)$ in (24) of Remark 1, it follows that

$$\begin{aligned} \|E_p(z)\|_2^2 &= \frac{1}{2\pi}\,\mathrm{trace}\left(\int_0^{2\pi} FT_p(e^{j\Omega})T_p^*(e^{j\Omega})F^T\,d\Omega\right) \\ &= \frac{1}{2\pi}\,\mathrm{trace}\left(\int_0^{2\pi} F\left[\frac{A^{p-1}}{e^{j\Omega(p-1)}}(e^{j\Omega}I - A + LC)^{-1}B\right] \times \left[\frac{A^{p-1}}{e^{-j\Omega(p-1)}}(e^{-j\Omega}I - A + LC)^{-1}B\right]^T F^T\,d\Omega\right) \\ &+ \frac{1}{2\pi}\,\mathrm{trace}\left(\int_0^{2\pi} F\left[\frac{A^{p-1}}{e^{j\Omega(p-1)}}(e^{j\Omega}I - A + LC)^{-1}B\right] \times \left[\sum_{i=0}^{p-2}\frac{A^i}{e^{-j\Omega(i+1)}}B\right]^T F^T\,d\Omega\right) \\ &+ \frac{1}{2\pi}\,\mathrm{trace}\left(\int_0^{2\pi} F\left[\sum_{i=0}^{p-2}\frac{A^i}{e^{j\Omega(i+1)}}\right] \times \left[\frac{A^{p-1}}{e^{-j\Omega(p-1)}}(e^{-j\Omega}I - A + LC)^{-1}B\right]^T F^T\,d\Omega\right) \\ &+ M(p), \end{aligned} \tag{39}$$

where

$$M(p) = \begin{cases} \displaystyle\frac{1}{2\pi}\,\mathrm{trace}\left(\int_0^{2\pi} F\left[\sum_{i=0}^{p-2}\frac{A^i}{e^{j\Omega(i+1)}}B\right] \times \left[\sum_{i=0}^{p-2}\frac{A^i}{e^{-j\Omega(i+1)}}B\right]^T F^T\,d\Omega\right) & p \ge 2 \\ 0 & p = 1 \end{cases}$$

Since the sum of the second integral and the third integral is identically equal to zero and the integrals of the cross terms in the fourth term also vanish

$$\|E_p(z)\|_2^2 = \frac{1}{2\pi}\,\mathrm{trace}\left\{FA^{p-1}\int_0^{2\pi}[(e^{j\Omega}I - A + LC)^{-1}BB^T \times (e^{-j\Omega}I - A + LC)^{-T}\,d\Omega]A^{(p-1)T}F^T\right\} + N(p) \tag{40}$$

where

$$N(p) = \begin{cases} \mathrm{trace}\left[F\left(\displaystyle\sum_{i=0}^{p-2} A^iBB^TA^{iT}\right)F^T\right] & p \ge 2 \\ 0 & p = 1 \end{cases}$$

For given plant model state-space matrices A and B, if the feedback gain matrix F is fixed, then the observer gain L is the only adjustable matrix which could change the H_2 norm of the error matrix. On the other hand, the covariance of the state estimation error in Proposition 3 can be written, according to the discrete-time Plancherel theorem (Francis, 1987) as follows

$$E\{e_{k|k-p}e^T_{k|k-p}\} = A^{p-1}\int_0^{2\pi}[(e^{j\Omega}I - A + LC)^{-1}BB^T \times (e^{-j\Omega}I - A + LC)^{-T}\,d\Omega]A^{(p-1)T} + S(p)$$

where

$$S(p) = \begin{cases} \displaystyle\sum_{i=0}^{p-2} A^iBB^TA^{iT}, & p \ge 2 \\ 0, & p = 1 \end{cases} \tag{41}$$

A comparison of the H_2 norm of the error matrix $E_p(z)$ in (40) with the trace of error covariance matrix $E\{e_{k|k-p}e^T_{k|k-p}\}$ in (41) reveals that any adjustment of L can only change the first term of both equations. Therefore, minimization of $\|E_p(z)\|_2$ is equivalent to that of $\mathrm{trace}[E\{e_{k|k-p}e^T_{k|k-p}\}]$ for $\forall p > 0$.

Next we proceed to find an optimal L that minimizes the trace $[E\{e_{k|k-p}e^T_{k|k-p}\}]$. It follows from Lemma 1, given below, that the minimum variance filter gain (with $p = 1$) also minimizes the trace$[E\{e_{k|k-p}e^T_{k|k-p}\}]$ while $p > 1$. Therefore, the optimal observer gain L that minimizes $\|E_p(z)\|_2$ is the same L that minimizes $\|E_1(z)\|_2$. According

to Lemma 2, the steady-state minimum variance gain with no measurement noise is the optimal gain.

Lemma 1 *for Proposition* 4. For the p-step predictor ($p \geq 1$), if the estimation error is defined as $e_{k|k-p} = x_k - \hat{x}_{k|k-p}$, then the filter gain L which minimizes the covariance $E\{e_{k|k-p}e^T_{k|k-p}\}$ is identical to the minimum variance filter gain.

Proof of Lemma 1. The proof directly follows the derivations in Chapter 5 of Maybeck (1979).

Lemma 2 *for Proposition* 4. For a fixed F, the H_2 optimization of one-step predictor error matrix given as

$$\min_L \|E_1(z)\|_2 = \min_L \|F(zI - A + LC)^{-1}B\|_2, \tag{42}$$

is the steady-state minimum variance filter gain where the plant and measurement noise covariance matrices, Q and R, are set to

$$Q = BB^T, \qquad R = \lim_{\rho \to 0} \rho I, \tag{43}$$

Proof of Lemma 2. The proof directly follows from the dual result of Theorem 3.1 in Yen and Horowitz (1989).

4. *Synthesis of the* p-*step delay compensator*

We propose a procedure for robust synthesis of the p-step delay compensator on the basis of the analytical results derived in Section 3. This two-stage procedure is structurally similar to the conventional LQR/LTR, and is described below.

- In the first stage, the target loop is designed assuming no delay (i.e. $p = 0$) and full-state feedback. The controller gain is optimized relative to a specified performance index. This is accomplished by shaping the target loop transfer matrix and the sensitivity matrix.
- In the second stage, according to Proposition 4, the observer gain L is made identically equal to the steady-state minimum variance filter gain. This gain is calculated by solving the steady-state Riccati equation for a fictitious measurement noise convariance matrix R which is set to ρI where ρ is a tunable scalar parameter and I is the identity matrix. The plant noise covariance matrix Q is set equal to BB^T. For a given ρ, the loop transfer matrix and the sensitivity matrix of the p-step delay compensator, derived in Proposition 2 and Remark 4, are computed. Finally, ρ is tuned to achieve the specified performance and robustness requirements. The similar procedure can be found in Maciejowski (1985) and Yen and Horowitz (1989).

The above synthesis procedure for p-step delay compensation has been verified by simulation of the flight control system of a fighter aircraft. The lumped delay of p sampling intervals is a representation of the sensor-to-controller and controller-to-actuator delays (that could result from a data communication network interconnecting specially distributed components of the flight control system). The (continuous time) plant model, linearized at the operating condition of 7.62 k and 0.9 Mach, is given below (Safonov *et al.*, 1981).

$$A = \begin{bmatrix} -0.0226 & -36.6170 & -18.8970 & 32.0900 & 3.2509 & -0.7626 \\ 0.0001 & -1.8997 & 0.9831 & -0.0007 & -0.1708 & -0.0050 \\ 0.0123 & 11.7200 & -2.6316 & 0.0009 & -31.6040 & 22.3960 \\ 0 & 0 & 1.000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -30.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & -30.0000 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 30 \\ 0 & 0 & 0 & 0 & 30 & 0 \end{bmatrix}^T, \quad C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

where the six plant state variables are forward speed, angle of attack, pitch rate, attitude angle, elevon actuator position, and canard actuator position, the two control inputs are elevon and canard signals, and the two output variables are angle of attack and attitude angle. The plant model was discretized at a sampling frequency of 1000 Hz which is sufficiently high relative to the desired operating frequency range of the closed loop system. On the basis of the discretized plant model, and a given set of performance and robustness specifications which can be found in Safonov *et al.* (1981), the control matrices for the target and delay-compensated systems were synthesized using a standard commerically available toolbox on a personal computer.

The full-state feedback regulator was designed using the standard LQR procedure with $Q_c = I_6$ and $R_c = 10^{-2}I$. The resulting optimal state feedback gain, F, is given below.

$$F = \begin{bmatrix} 7.3570 & -19.395 & -9.9957 & -15.975 & 8.8425 & 0.70813 \\ -4.2423 & 11.685 & 6.6904 & 10.395 & -0.70716 & 8.2537 \end{bmatrix}$$

For tuning the filter gain in the LTR design procedure, the plant noise covariance was set as $Q = BB^T$ and the (fictitious) measurement noise covariance was set to $R = \rho I$, where ρ was tuned in the range of 10^{-4}–10^{-1}. Initially a series of simulation experiments were conducted with no delays, i.e. $p = 0$, to verify that the system performance is degraded as p is increased and the error recovery becomes impossible. The LTR procedure yielded good results for recovering the full-state feedback robustness properties at $p = 0$ as expected. The system performance improved as the measurement noise (i.e. ρ) was reduced and the loop sensitivity matrix converged to the target sensitivity matrix as ρ was made to approach zero.

In the simulation experiments, the scalar parameter ρ was tuned to adjust the observer gain L for both one-step and two-step compensators such that the stability robustness for each case (i.e. $p = 1$ and $p = 2$) bears a desired safe margin relative to the target system while the state feedback gain is retained at the optimal value F for the target system. The robustness margin in the frequency range of 10–1000 Hz was set to 10 dB in terms of the maximum singular value of the loop transfer matrix to overcome the detrimental effects of loss of phase margin resulting from delays. The parameter ρ in the design procedure of the delay compensator was adjusted to satisfy this requirement for robustness. The respective values of ρ and the resulting filter gain L that were used for one-step and two-step delay compensators are given in Table 1.

Figure 4 shows a comparison of the maximum and minimum singular values of the loop transfer matrices for the one-step delay compensated (i.e. $p = 1$) and two-step delay-compensated (i.e. $p = 2$) systems. (Note: The target loop with full-state feedback and no delay merely serves as a reference for the synthesis procedure.) The minimum singular values of the loop transfer matrices of the individual systems represent the lower bounds of their respective performance, and their maximum singular values represent the upper bounds of stability robustness. As stated earlier, the observer gain matrices of both compensators are adjusted by tuning ρ such that their maximum singular values are about 10 dB lower than that of the target system in the range of 10–1000 Hz. The minimum singular values for both compensators, as seen in Fig. 4, are significantly lower than that for the target system. This is expected in view of the reduction in observer gain resulting in decreased loop transfer gain. However, the two-step compensator consistently exhibits a lower performance (of about 12 dB) relative to the one-step compensator because of lower observer gain due to a higher value of the parameter ρ.

One major criterion in the synthesis procedure is to reduce the difference between the sensitivity matrices of the target loop and the delay compensated loop (in the H_2 sense) as

TABLE 1. OBSERVER GAINS AND ADJUSTING PARAMETERS OF THE p-STEP DELAY COMPENSATOR

One step delay compensator $\rho = 1 \times 10^{-12}$

$$L = \begin{pmatrix} -0.42 \times 10^{7} & 0.2 \times 10^{1} & 0.33 \times 10^{3} & 0.000 & -0.6252 \times 10^{4} & 0.341 \times 10^{3} \\ -0.41 \times 10^{7} & 0.000 & 0.772 \times 10^{3} & 0.1 \times 10^{3} & 0.2703 \times 10^{4} & 1.217 \times 10^{4} \end{pmatrix}^{T}$$

Two step delay compensator $\rho = 5 \times 10^{-8}$

$$L = \begin{pmatrix} -6.6239 & 3.4811 \times 10^{-1} & 3.7994 \times 10 & 1.2893 \times 10^{-1} & -1.3923 \times 10^{2} & 1.5987 \times 10 \\ -4.1893 & 1.1793 \times 10^{-1} & 3.0325 \times 10 & 2.2847 \times 10^{-1} & 1.9327 \times 10 & 1.2122 \times 10^{2} \end{pmatrix}^{T}$$

much as possible. From this perspective, Fig. 5 shows comparisons of the maximum and minimum singular values of the sensitivity matrices for the target, one-step delay compensated (i.e. $p = 1$) and two-step delay compensated (i.e. $p = 2$) systems. As shown in Table 1, the observer gains of the delay compensators are tuned to their respective values such that the specification for stability robustness is satisfied. This renders the minimum singular values of the sensitivity matrices of the two compensators remain close to each other and above that of the target system except in the high frequency region where all of them move towards 0 dB. Therefore, the sensitivity of the delay compensators reduces at high frequency, which is a very desirable feature from the point of view of stability robustness. On the other hand, the performance of the delay compensator (in the low frequency range) degrades as p is increased. It follows from Fig. 5 that $\sigma(L_1(e^{j\omega})) - 5.5$ dB and $\sigma(L(e^{j\omega})) - 31.8$ dB in the frequency range of 10–10 Hz.

5. Conclusions

In many real time distributed control systems such as advanced aircraft, spacecraft and autonomous manufacturing plants, the sensor and control signals within a feedback loop are subjected to delays induced by multiplexed data communication networks or due to priority interruption at the control computer. From this perspective, a procedure has been developed for robust compensation of delays in multi input–multi output discrete time control systems. This control synthesis procedure is an extension of the standard loop transfer recovery (LTR) from one step prediction to the general case of p-step prediction ($p > 1$) and is carried out in two steps: (1) evaluation of the state feedback gain assuming full state feedback and no induced delays, and (2) tuning of the filter gain by varying a scalar parameter representing the (fictitious) measurement noise covariance matrix. The delay compensated system, albeit inferior in performance relative to the non delayed full state feedback system, can be synthesized for a given value of p. The synthesis procedure is demonstrated via simulation of the flight control system of a fighter aircraft.

The major conclusion derived from the analytical work reported in this paper is as follows. The concept of the steady state minimum variance filter gain as the H minimization solution of the difference between the target sensitivity matrix and the actual sensitivity matrix for one step prediction does not hold for p-step prediction ($p > 1$). The conclusions from the perspective of control synthesis are: (i) it is impossible to tune the observer gain for a delayed system (i.e. $p > 1$) to fully recover the target loop characteristics, and (ii) if the delay compensated system is designed to satisfy a specified requirement of stability robustness, then its performance decreases as p increases. The results and conclusions are also applicable if the plant model has inherent delays which have the same effects on the loop recovery error.

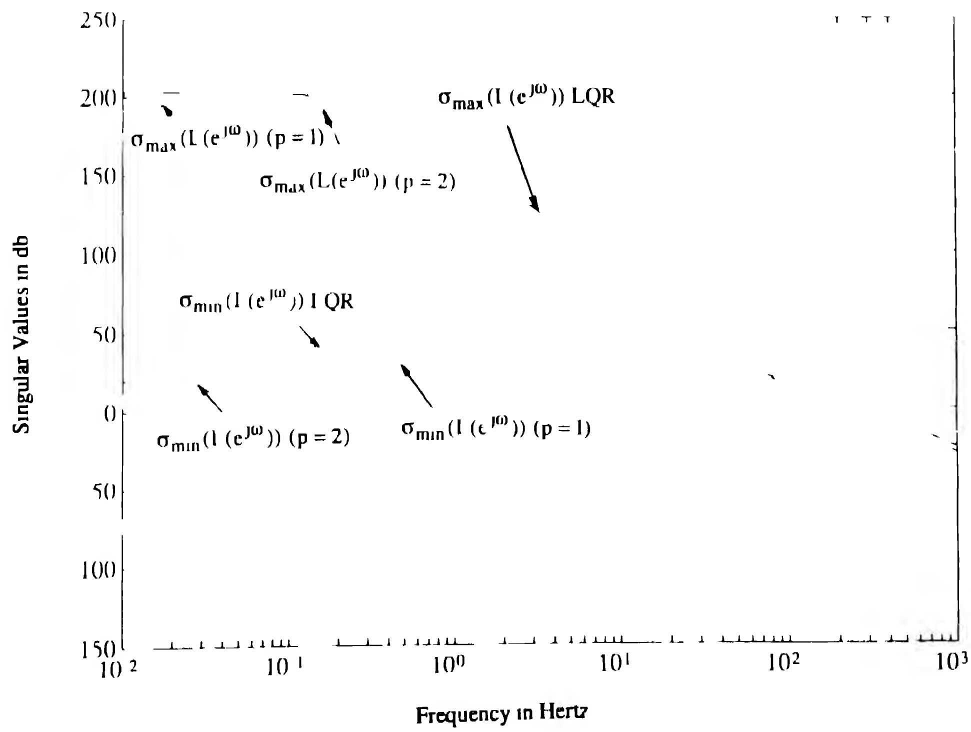


FIG. 4. Comparison loop transfer matrices.

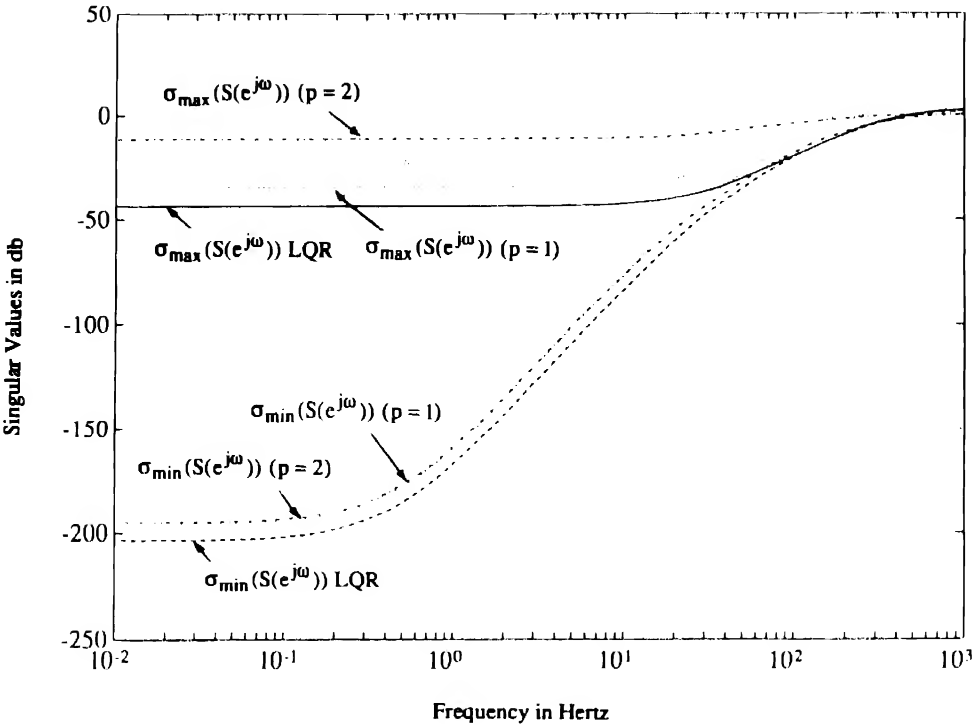


FIG. 5. Comparison of sensitivity matrices.

Acknowledgement—The authors acknowledge the benefits of discussion with Dr Zhihong Zhang of General Motors Systems Engineering Center, Troy, MI.

References

Belle Isle, A. P. (1975). Stability of systems with nonlinear feedback through randomly time-varying delays. *IEEE Trans. Aut. Control*, **AC-20**, 67–75.

Churchill, R. V. (1960). *Complex Variables and Applications*. McGraw Hill, New York.

Doyle, J. C. and G. Stein (1979). Robustness with observers. *IEEE Trans. on Aut. Control*, **AC-24**, 607–611.

Doyle, J. C. and G. Stein (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. on Aut. Control*, **AC-26**, 4–16.

Francis, B. A. (1987). *A Course in H_∞ Control Theory*. Springer-Verlag, New York.

Halevi, Y. and A. Ray (1988). Integrated communication and control systems: Part I—analysis. *ASME J. of Dynamic Systems, Measurements and Control*, pp. 367–373.

Ishihara, T. (1988). Sensitivity properties of a class of discrete-time LQG controllers with computation delays. *Systems and Control Letters*, **11**, 299–307.

Kinnaert, M. and Y. Peng (1990). Discrete-time LQG/LTR technique for systems with time delays. *Systems and Control Letters*, **15**, 303–311.

Liou, L.-W and A. Ray (1990). Integrated communication and control systems: Part III—non-identical sensor and controller sampling. *ASME J. of Dynamic Systems, Measurement and Control*, pp. 364–371.

Luck, R. and A. Ray (1990). An observer-based compensator for distributed delays. *Automatica*, **26**, 903–908.

Maciejowski, J. M. (1985). Asymptotic recovery for discrete-time systems. *IEEE Trans. on Aut. Control*, **AC-30**, 602–605.

Maybeck, P. S. (1979). *Stochastic Models, Estimation, and Control, Volume* 1. Academic Press, San Diego.

Ray, A. and Y. Halevi (1988). Integrated communication and control systems; Part II—design considerations. *ASME J. of Dynamic Systems, Mesasurement and Control*, pp. 374–381.

Safonov, M. G., A. J. Laub and G. L. Hartmann (1981). Feedback properties of multivariable systems: the role and use of the return difference matrix. *IEEE Trans. Aut. Control*, **AC-26**, 47–65.

Shaked, U. (1985) Explicit solution to the singular discrete-time stationary linear filtering problem. *IEEE Trans. on Aut. Control*, **AC-30**, 34–47.

Stein, G. and M. Athans (1987). The LQG/LTR procedure for multivariable feedback control design. *IEEE Trans. on Aut. Control*, **AC-32**, 105–114.

Yen, J. and R. Horowitz (1989). Discrete time H_2 loop transfer recovery. *Proc. of ASME Winter Annual Meeting*, ASME Paper No. 89-WA/DSC-35.

Zhang, Z. and J. S. Freudenberg (1991). On discrete-time loop transfer recovery. *Proc. of* 1991 *American Control Conf.*, pp. 2214–2219.

Automatica, Vol 29, No 2, pp 439 444, 1993
Printed in Great Britain

0005-1098/93 $6 00 + 0.00

Brief Paper

Heuristically Enhanced Feedback Control of Constrained Systems: The Minimum Time Case*

M. SZNAIER† and M. J. DAMBORG‡

Key Words—Computer control; constrained systems; dynamic programming; multivariable control systems; stability; suboptimal control.

Abstract—Recent advances in computer technology have spurred new interest in the use of feedback controllers based upon the use of on-line optimization Still, the use of computers in the feedback loop has been hampered by the limited amount of time available for computations In this paper we propose a feedback controller based upon the use of on-line contrained optimization in the feedback loop. The optimization problem is simplified by making use of the special structure of time-optimal systems, resulting in a substantial dimensionality reduction. These results are used to show that the proposed controller yields asymptotically stable systems, provided that enough computation power is available to solve on-line a constrainted optimization problem considerably simpler than the original.

1. *Introduction*

A SUBSTANTIAL NUMBER of control problems can be summarized as the problem of designing a controller capable of achieving acceptable performance under design constraints. This statement looks deceptively simple, but even in the case where the system under consideration is linear time-invariant, the problem is far from solved.

During the last decade, substantial progress has been achieved in the design of linear controllers. By using a parametrization of all internally stabilizing linear controllers in terms of a stable transfer matrix Q, the problem of finding the "best" linear controller can be formulated as an optimization problem over the set of suitable Q (Boyd *et al.*, 1988). In this formulation, additional specifications can be imposed by further constraining the problem. However, most of these methods can address time-domain constraints only in a conservative fashion. Hence if the constraints are tight this approach may fail to find a solution, even if the problem is feasible (in the sense of having a, perhaps non-linear, solution).

Classically, control engineers have dealt with time-domain constraints by allowing inputs to saturate, in the case of actuator constraints, and by switching to a controller that attempts to move the system away from saturating constraints, in the case of state constraints. Although these methods are relatively simple to use, they have several serious shortcomings, perhaps the most important being their inability to handle constraints in a general way. Hence, they require *ad hoc* tuning of several parameters making extensive use of simulations.

Alternatively, the problem can be stated as an optimization problem (Frankena and Sivan, 1979). Then, mathematical programming techniques can be used to find a solution (see for instance Zadeh and Whalen, 1962; Fegley *et al.*, 1971, and references therein). However, in most cases the control law generated is an open-loop control that has to be recalculated entirely, with a considerable computational effort, if the system is disturbed. Conceivably, the set of open loop control laws could be used to generate a closed loop control law by computing and storing a complicated field of extremals (Judd *et al.*, 1987). However, this alternative requires extensive amounts of off-line computation and of storage.

Because of the difficulties with the optimal control approach, other design techniques, based upon using a Lyapunov function to design a stabilizing controller, have been suggested (Gutman and Hagander, 1985). However, these techniques tend to be unnecessarily conservative. Moreover, several steps of the design procedure involve an extensive trial and error process, without guarantee of success (see example 5.3 in Gutman and Hagander, 1985).

Recently, several techniques that exploit the concept of maximally invariant sets to obtain static (Gutman and Cwikel, 1986; Vassilaki *et al.*, 1988; Benzaouia and Burgat, 1988; Bitsoris and Vassilaki, 1990; Blanchini, 1990; Sznaier, 1990; Sznaier and Sideris, 1991a) and dynamic (Sznaier and Sideris, 1991b; Sznaier, 1991) linear feedback controllers have been proposed. These controllers are particularly attractive due to their simplicity. However, it is clear that only a fraction of the feasible constrained problems admit a linear solution. Furthermore, performance considerations usually require the control vector to be on a constraint boundary and this clearly necessitates a non-linear controller capable of saturating.

Finally, in the last few years, there has been a renewed interest in the use of feedback controllers based upon the use of on-line minimization. Although this idea was initially proposed as far back as 1964 (Dreyfus, 1964), its implementation has become possible only during the last few years, when the advances in computer technology made feasible the solution of realistically sized optimization problems in the limited time available. Consequently, theoretical results concerning the properties of the resulting closed-loop systems have started to emerge only recently. In Sznaier (1989) and Sznaier and Damborg (1990) we presented a theoretical framework to analyse the effects of using on-line optimization and we proposed a controller guaranteed to yield asymptotically stable systems. However, although these theoretical results represent a substantial advance over some previously used *ad hoc* techniques, in some cases they are overly conservative, requiring the on-line solution of a large optimization problem. Since in most sampled control systems the amount of time available

* Received 29 April 1991; revised 24 December 1991; revised 17 May 1992; received in final form 4 June 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor E. K. Kreindler under the direction of Editor H. Kwakernaak.

† Author to whom all correspondence should be addressed. Department of Electrical Engineering, University of Central Florida, Orlando, FL 32816, U.S A.

‡ Department of Electrical Engineering, FT-10, University of Washington, Seattle, WA 98195, U.S.A.

between samples is very limited, this may preclude the use of the proposed controller in many applications.

In this paper we present a suboptimal feedback controller for the minimum-time control of discrete time constrained systems. Following the approach presented in Sznaier and Damborg (1990) this controller is based upon the solution, during the sampling interval, of a sequence of optimization problems. We show that by making use of the special structure of time-optimal systems the proposed algorithm results in a significant reduction of the dimensionality of the optimization problem that must be solved on-line, hence allowing for the implementation of the controller for realistically sized problems.

The paper is organized as follows. In Section 2 we introduce several required concepts and we present a formal definition to our problem. In Section 3 we present the proposed feedback controller and the supporting theoretical results. The main result of this section shows that by relinquishing theoretical optimality we can find a stabilizing suboptimal controller that allows for a substantial reduction of the dimensionality of the optimization problem that must be solved on-line. Finally, in Section 4, we summarize our results and we indicate directions for future research.

2. *Problem formulation and preliminary results*

2.1. *Statement of the problem.* Consider the linear, time invariant, controllable discrete time systems modeled by the different equation

$$x_{k+1} = Ax_k + Bu_k, \quad k = 0, 1, \ldots \tag{S}$$

with initial condition x_0 and the constraints

$$\begin{gathered} u_k \in \Omega \subset R^m \quad x_k \in \mathcal{G} \subset R^n \\ \mathcal{G} = \{x : |Gx| \le \gamma\}, \quad \Omega = \{u : |Wu| \le \omega\} \end{gathered} \tag{C}$$

where $\gamma \in R^p$, $\omega \in R^q$, $\gamma_i, \omega_i > 0$, $G \in R^{p \cdot n}$, $W \in R^{q \cdot m}$ with full column rank, x, indicates x is a vector quantity and where the inequalities (C) should be interpreted in a component by component sense. Furthermore, assume as usual that A^{-1} exists. Our objective is to find a sequence of admissible controls, $u_k[x_k]$, that minimizes the transit time to the origin. Throughout the paper we will refer to this optimization problem as problem (P) and we will assume that it is well posed in the sense of having a solution. In Section 3 we give a sufficient condition on $\mathcal{G}$ for (P) to be feasible.

2.2. *Definitions and preliminary results.* In order to analyse the proposed controller we need to introduce some definitions and background theoretical results. We begin by formalizing the concept of null controllable domain and by introducing a constraint-induced norm.

Definition 1. The Null Controllable domain of (S) is the set of all points $x \in \mathcal{G} \subset R^n$ that can be steered to the origin by applying a sequence of admissible controls $u_k \in \Omega \subset R^m$ such that $x_k \in \mathcal{G}$, $k = 0, 1, \ldots$. The Null Controllable domain of (S) will be denoted as C_∞. The Null Controllable domain in j or fewer steps will be denoted as $C_j \subseteq C_\infty$.

Definition 2. The Minkowsky Functional (or gauge) p of a convex set $\mathcal{G}$ containing the origin in its interior is defined by

$$p(x) = \inf_{r>0} \left\{ r : \frac{x}{r} \in \mathcal{G} \right\}$$

A well-known result in functional analysis (see for instance Conway, 1990) establishes that p defines a seminorm in R^n. Furthermore, when $\mathcal{G}$ is balanced and compact, as in our case, the seminorm becomes a norm. In the sequel we will denote this norm as $p(x) = \|\Gamma^{-1}Gx\|_\infty \triangleq \|x\|_{\mathcal{G}}$, where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_p)$.

Remark 1. The set $\mathcal{G}$ can be characterized as the unity ball in $\|\ \|_{\mathcal{G}}$. Hence, a point $x \in \mathcal{G}$ iff $\|x\|_{\mathcal{G}} \le 1$.

Next, we formalize the concept of underestimate of the cost to go. We will use this concept to determine a sequence of approximations that converges to the solution of (P).

Definition 3. Let O be a convex open set containing the origin and such that for all the optimal trajectories starting out in O the constraints (C) are not effective, and let $J_0(x)$ be the optimal cost-to-go from the state x. A function $g: R^n \to R$ such that

$$0 \le g(x) \le J_0(x)\ \forall x \in \mathcal{G},$$

$$g(x) = J_0(x)\ \forall x \in O$$

will be called an underestimate of the cost-to-go relative to the set O.

The following theorem, where we show that problem (P) can be exactly solved by solving a sequence of suitable approximations, provides the theoretical motivation for the proposed controller.

Theorem 1. Let O be the set introduced in Definition 3 and let $x_k^u(\xi)$ be the (unconstrained) optimal trajectory corresponding to the initial condition $\xi \in O$. Finally let $g(x): R^n \to R$ be an underestimate relative to O. Consider the following optimization problems

$$\min_{\mathcal{U}} \{J(x) = N\} = \min_{\mathcal{U}} \left\{ \sum_{k=1}^{N} 1 \right\} \tag{1}$$

$$\min_{\mathcal{U}} \{J_m(x) = m - 1 + g(x_m)\} = \min_{\mathcal{U}} \left\{ \sum_{k=1}^{m-1} 1 + g(x_m) \right\} \quad m \le N \tag{2}$$

subject to (C), where $\mathcal{U} = \{u_0, u_1, \ldots\}$. Then an optimal trajectory x_k^0, $k = 1, 2, \ldots, m$ which solves (2) extended by defining $x_k^0 = x_k^u(x_m^0)$, $k = m+1, \ldots, N$ is also a solution of (1) provided that $x_m^0 \in O$.

Proof. The proof follows by noting that the theorem corresponds to a special case of Theorem 1 in Sznaier and Damborg (1990) with $L_k(x_k, u_k) = 1$. ◇

It follows that problem (P) can be exactly solved by using the sampling interval to solve a sequence of optimization problems of the form (2), with increasing m, until a number m_0 and a trajectory x_k such that $x_{m_0} \in O$ are obtained. However this approach presents the difficulty that the asymptotic stability of the resulting closed loop system can not be guaranteed when there is not enough time to reach the region O.

In our previous work (Sznaier, 1989; Sznaier and Damborg, 1990) we solved this difficulty by imposing an additional constraint (which does not affect feasibility) and by using an optimization procedure based upon the quantization of the control space. By quantizing the control space, the attainable domain from the initial condition can be represented as a tree with each node corresponding to one of the attainable states. Thus the original optimal control problem is recast as a tree problem that can be efficiently solved using heuristic search techniques based upon an underestimate of the cost to go (Winston, 1984). We successfully applied this idea to minimum time and quadratic regulator problems. However, as we noted there, in some cases the results, based upon a worst-case analysis, proved to be overly conservative. As a result the optimization problem quickly became untractable. This phenomenon is illustrated in the following example.

2.3. *A realistic problem.* Consider the minimum time control of an F-100 jet engine. The system at intermediate power, sea level static and PLA = 83° can be represented by

Sznaier (1989):

$$A = \begin{pmatrix} 0.8907 & 0.0474 & -0.0980 & 0.2616 & 0.0689 \\ 0.0237 & 0.9022 & -0.0202 & 0.1057 & 0.0311 \\ 0.0233 & -0.0149 & 0.8167 & 0.2255 & 0.0295 \\ 0.0 & 0.0 & 0.0 & 0.7788 & 0.0 \\ -0.0979 & 0.3532 & 0.3662 & 0.6489 & 0.0295 \end{pmatrix},$$

$$\gamma = \begin{pmatrix} 50.0 \\ 64.0 \\ 20.0 \\ 5.0 \\ 18.1 \end{pmatrix}$$

$$B = \begin{pmatrix} 0.0213 & -0.3704 \\ 0.0731 & -0.1973 \\ -0.0357 & -0.5538 \\ 0.2212 & 0.0 \\ 0.0527 & -3.9068 \end{pmatrix}$$

$$G = I; \quad \Omega = \{u \in R^2 : |u_1| \leq 31.0; \; |u_2| \leq 200.0\}. \tag{3}$$

The sampling time for this system is 25 msec. In this case equation (21) in Sznaier and Damborg (1990) yields $\sim 10^3$ nodes for each level of the tree, which clearly precludes the real-time implementation of the algorithm proposed there.

3. *Proposed control algorithm*

In this section we indicate how the special structure of time-optimal systems can be used to reduce the dimensionality of the optimization problem that must be solved on-line. Specifically, we use a modification of the Discrete Time Minimum Principle to show that the points that satisfy a necessary condition for optimality are the corners of a subset of Ω. Hence, only these points need to be considered by the optimization algorithm.

3.1. *The modified discrete time minimum principle.* We begin by extending the Local Discrete Minimum Principle (Butkovskii, 1963) to Frechet differentiable terminal-cost functions and constraints of the form (C). Note that in its original form, the minimum principle requires the state constraint set $\mathscr{G}$ to be open, while in our framework it is compact.

Theorem 2. Consider the problem (P') defined as

$$\min_{u_k \in \Omega} S(x_N), \tag{4}$$

subject to:

$$x_{k+1} = Ax_k + Bu_k \triangleq f(x_k, u_k), \quad x_0, N \text{ given} \tag{5}$$

$$\|x_{k+1}\|_{\mathscr{G}} < \|x_k\|_{\mathscr{G}}, \tag{6}$$

where S is Frechet differentiable. Let the co-states ψ_k be defined by the difference equation:

$$\psi_k' = \psi_{k+1}' \frac{\partial f(x_{k+1}, u_{k+1})}{\partial x'} = \psi_{k+1}' A, \qquad \psi_{N-1}' = \frac{\partial S(x_N)}{\partial x'}. \tag{7}$$

Finally, define the Hamiltonian as:

$$H(x_k, u_k, \psi_k) = \psi_k' f(x_k, u_k). \tag{8}$$

Then, if:

$$\max_{\|x\|_{\mathscr{G}}=1} \left\{ \min_{u \in \Omega} \|Ax + Bu\|_{\mathscr{G}} \right\} < 1, \tag{9}$$

the following results hold; (i) problem (P') is feasible; (ii) for any initial condition $x_0 \in \mathscr{G}$ the resulting trajectory $\{x_k\}$ is admissible; and (iii) a necessary condition for optimality is:

$$H(x_k^*, u_k^*, \psi_k^*) = \min_{u \in O_u \subseteq \Omega_1} H(x_k^*, u, \psi_k^*), \quad k = 1, \ldots, N-1, \tag{10}$$

where

$$\Omega_1(x_k, k) = \{u \in \Omega : \|x_{k+1}\|_{\mathscr{G}} \leq (1-\epsilon)\|x_k\|_{\mathscr{G}}\}, \tag{11}$$

where O_u is some neighborhood of u, $\epsilon > 0$ is chosen such that Ω_1 is not empty and where * denotes the optimal trajectory.

Proof. Feasibility follows from (9) and Theorem 3.1 in Gutman and Cwikel (1986) (or as a special case of Theorem 2 in Sznaier and Damborg (1990)). Since $x_0 \in \mathscr{G}$, $\|x_0\|_{\mathscr{G}} \leq 1$. From (9) $\|x_k\|_{\mathscr{G}} < 1$ and therefore $x_k \in \mathscr{G}$ for all k. To prove (iii) we proceed by induction. From (6) it follows that there exists $\epsilon > 0$ such that Ω_1 is not empty. From the definition of Ω_1 it follows that for any $u_k \in \Omega_1(x_k, k)$ there exists a neighborhood $O_u \subseteq \Omega_1$, not necessarily open, where (6) holds. Hence, if $x_k \in \mathscr{G}$, $x_{k+1} = f(x_k, u_k) \in \mathscr{G}\, \forall u_k \in O_u$. Let $\bar{x}_k$ denote a non-optimal feasible trajectory obtained by employing the non-optimal control law $\bar{u}_{N-1}$ at stage $k = N-1$. Consider a neighborhood $O_u \subseteq \Omega$ of u_{N-1}^* such that the state constraints are satisfied for all the trajectories generated employing controls in O_u. For any such trajectory $\bar{x}$, $\bar{u}$ we have:

$$S(x_N^*) \leq S(\bar{x}_N). \tag{12}$$

Hence:

$$\frac{\partial S}{\partial x'} \Delta x_N = \frac{\partial S}{\partial x'}\Big|_{x_N^*} (f(x_{N-1}^*, \bar{u}_{N-1}) - (f(x_{N-1}^*, u_{N-1}^*)) \geq 0, \tag{13}$$

$$\begin{aligned} H(x_{N-1}^*, \bar{u}_{N-1}, \psi_{N-1}^*) &= \frac{\partial S}{\partial x'} f(x_{N-1}^*, \bar{u}_{N-1}) \\ &\geq \frac{\partial S}{\partial x'}\Big|_{x_N^*} f(x_{N-1}^*, u_{N-1}^*) \\ &= H(x_{N-1}^*, u_{N-1}^*, \psi_{N-1}^*). \end{aligned} \tag{14}$$

Consider now a neighborhood $O_u \subseteq \Omega$ of u_k^* such that all the trajectories obtained by replacing u_k^* by any other control in O_u satisfy the constraints and assume that (10) does not hold for some $k < N-1$. Then, there exists at least one trajectory $\bar{x}$, $\bar{u}$ such that:

$$H(x_k^*, \bar{u}_k, \psi_k^*) < H(x_k^*, u_k^*, \psi_k^*). \tag{15}$$

Therefore:

$$0 > \psi_k^{*\prime}(f(x_k^*, \bar{u}_k) - f(x_k^*, u_k^*)) = \psi_k^{*\prime} \Delta x_{k+1}. \tag{16}$$

Hence:

$$\begin{aligned} H(\bar{x}_{k+1}, u_{k+1}^*, \psi_{k+1}^*) &- H(x_{k+1}^*, u_{k+1}^*, \psi_{k+1}^*) \\ &= \psi_{k+1}^{*\prime} f(\bar{x}_{k+1}, u_{k+1}^*) - \psi_{k+1}^{*\prime} f(x_{k+1}^*, u_{k+1}^*) \\ &= \psi_{k+1}^{*\prime}\left(\frac{\partial f}{\partial x'}\Big|_{x_{k+1}^*}\right) \Delta x_{k+1} < 0, \end{aligned} \tag{17}$$

or:

$$H(\bar{x}_{k+1}, u_{k+1}^*, \psi_{k+1}^*) < H(x_{k+1}^*, u_{k+1}^*, \psi_{k+1}^*). \tag{18}$$

Using the same reasoning we have:

$$\begin{aligned} H(\bar{x}_{k+2}, u_{k+2}^*, \psi_{k+2}^*) &< H(x_{k+2}^*, u_{k+2}^*, \psi_{k+2}^*), \\ &\vdots \\ H(\bar{x}_{N-1}, u_{N-1}^*, \psi_{N-1}^*) &< H(x_{N-1}^*, u_{N-1}^*, \psi_{N-1}^*). \end{aligned} \tag{19}$$

From (19) it follows that

$$0 > \psi_{N-1}^{*\prime} \Delta x_N = \frac{\partial S}{\partial x'} \Delta x_N = \Delta S, \tag{20}$$

against the hypothesis that $S(x_N^*)$ was a minimum. ◇

3.2. *Using the modified discrete minimum principle.* In this section we indicate how to use the results of Theorem 2 to

generate a set of points that satisfy the necessary conditions for optimality. In principle, we could apply the discrete minimum principle to problem (P') by taking $S(x_N) = \|x_N\|_2^2$ and solving a sequence of problems, with increasing N, until a trajectory x^* and a number N_0 such that $x^*_{N_0} = 0$ are found. However note that Theorem 2 does not add any information to the problem since;

$$\psi^{*\prime}_{N-1} = \frac{\partial S(x^*_N)}{\partial x'} = x|_0 = 0. \tag{21}$$

It follows that $\psi_k = 0\,\forall k$, and hence the optimal trajectory corresponds to a "singular arc". Therefore, nothing can be inferred *a priori* about the controls. In order to be able to use the special structure of the problem, we would like the co-states, ψ, to be non-zero.

Consider now the special case of problem (P') where $S(x_N) = \frac{1}{2}\|x_N\|_2^2$ (with fixed terminal time N). Let n be the dimension of the system (S) and assume that the initial condition x_0 is such that the origin cannot be reached in N stages. Then, from (7):

$$\begin{aligned} \psi^{*\prime}_k &= \psi^{*\prime}_{k+1} A, \\ \psi^{*\prime}_{N-1} &= x_N \neq 0. \end{aligned} \tag{22}$$

It follows (since A is regular) that $\psi^*_k \neq 0\,\forall k$. Furthermore, since (S) is controllable, C_n (null controllability region in n steps) has dimension n (Sznaier, 1989). It follows that, by taking N large enough $x_N \in C_n$. Hence an approximate solution to (P) can be found by solving (P') for N such that $x_N \in C_n$ and by using Linear Programming to find the optimal trajectory from x_N to the origin.

Theorem 3. The optimal control sequence $\mathcal{U} = \{u^*_0 \cdots u^*_{N-1}\}$ that solves problem (P') is always in the boundary of the set Ω_1. Further, the control sequence can always be selected to be a corner point of such a set.

Proof. Since the constraints are linear and $\psi^*_k \neq 0$, it follows that the control u^*_k that solves (10) belongs to the boundary of the set $\Omega_1(x_k, k)$. Further, except in the case of degeneracies, i.e. when ψ^*_k is parallel to one of the boundaries of Ω_1, the control u^*_k must be a corner point of the set. In the case of degeneracies, all the points of the boundary parallel to the co-state yield the same value of the Hamiltonian and therefore the optimal control u_k still can be selected to be a corner of Ω_1. ◇

3.3. *Algorithm* H_{MP}. In this section we apply the results of Theorem 3 to obtain a suboptimal stabilizing feedback control law. From Theorem 3 it follows that problem (P') can be solved by using the following algorithm.

Algorithm H_{MP} (Heuristically Enhanced Control using the minimum principle)

(1) Determine ϵ for equation (11) (for instance using Linear Programming off-line). Let $O = C_1$, and determine an underestimate $g(\cdot)$ relative to O.
(2) Let x_k be the current state of the system:
  (a) If $x_k \in C_n$, null controllability region in n steps, solve problem (P) exactly using Linear Programming.
  (b) If $x_k \notin C_n$ generate a tree by considering all possible trajectories starting at x_k with controls that lie in the corners of the polytope $\Omega_1(x_k, k)$. Search the tree for a minimum cost trajectory to the origin, using heuristic search algorithms and $g(\cdot)$ as heuristics.
  (c) If there is no more computation time available for searching and the region O has not been reached, use the minimum partial cost trajectory that has been found.
(3) Repeat step 2 until the region the origin is reached.

Remark 2. Note that by solving problem (P') instead of (P) we are relinquishing optimality, strictly speaking, since the trajectory that brings the system closer to C_n is not necessarily the trajectory that will yield minimum transit time to the origin. However, for any "reasonable" problem, we would expect both trajectories to be close in the sense of yielding approximately equal transit times (in the next section we will provide an example where this expectation is met).

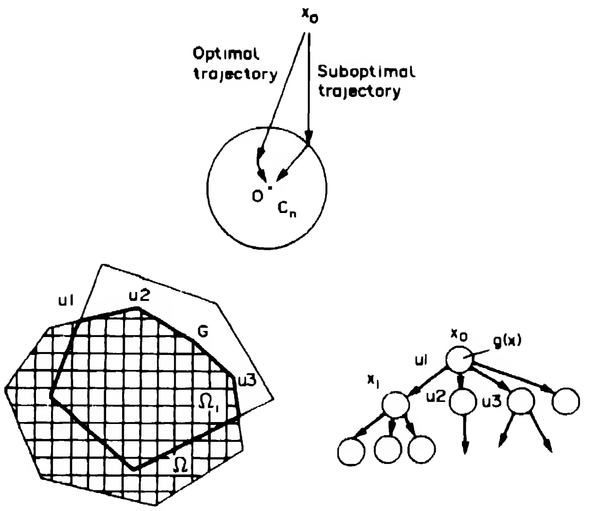


FIG. 1. Using the discrete minimum principle to limit the search.

Remark 3. Algorithm H_{MP} considers at each level of the tree only the control that lie in the corners of Ω_1, as illustrated in Fig. 1, therefore presenting a significant reduction of the dimensionality of the problem. However, the algorithm requires finding the vertices of a polytope given by a set of inequalities in R^m and this is a non-trivial computational geometry problem.

Theorem 4. The closed loop system resulting from the application of algorithm H_{MP} to problem (P') is asymptotically stable, provided that there is enough computational power available to search one level of the tree during the sampling interval.

Proof. From Theorem 2 it follows that $\|x_k\|_\mathscr{G}$ is monotonically decreasing in $\mathscr{G}$. Hence the system is guaranteed to reach the region C_n. But, since the solution to (P) is known in this region, it follows that the exact cost-to-go is a Lyapunov function for the system in C_n. Thus the resulting closed-loop system is asymptotically stable. ◇

3.4. *The heuristic for algorithm* H_{MP}. In order to complete the description of algorithm H_{MP} we need to provide a suitable underestimate $g(x)$. In principle, an estimate of the number of stages necessary to reach the origin can be found based upon the singular value decomposition of the matrices A and B, using the same technique that we used in Sznaier and Damborg (1990). However, in many cases of practical interest such as the F-100 jet engine of Section 2.3, the limitation in the problem is essentially given by the state constraints (i.e. the control authority is large). In this situation, this estimate yields an unrealistically low value for the transit time, resulting in poor performance.

The performance of the algorithm can be improved by considering an heuristic based upon experimental results. Recall that optimality depends on having, at each time interval, an underestimate $g(x)$ of the cost-to-go. Consider now the Null Controllability regions (C_k). It is clear that if they can be found and stored, the true transit time to the origin is known. If the regions are not known but a supraestimate C^s_k such that $C_k \subsetneq C^s_k$ is available, a suitable underestimate $g(x)$, can be obtained by finding the largest k such that $x \in C^s_k$ and $x \notin C^s_{k-1}$. However, in general these supraestimates are difficult to find and characterize (Sznaier, 1989). Hence, it is desirable to use a different heuristic, which does not require the use of these regions. From the convexity of Ω and $\mathscr{G}$ it follows that the regions C_k

are convex. Therefore, a subestimate C_{sk} such that $C_{sk} \subseteq C_k$ can be found by finding points in the region C_k and taking C_{sk} as their convex hull. Once a subestimate of C_k is available, an estimate $\bar{g}(x)$ of the cost-to-go can be found by finding the largest k such that $x \in C_{sk}$ and $x \notin C_{sk-1}$. Note that this estimate is not an underestimate in the sense of Definition 3. Since $C_{sk} \subseteq C_k$ then $x_k \in C_k \not\Rightarrow x_k \in C_{sk}$ and therefore $\bar{g}(x_k)$ is not necessarily $\leq k$. Thus, Theorem 1 that guarantees that once the set O has been reached the true optimal trajectory has been found is no longer valid. However, if enough points of each region are considered so that the subestimates are close to the true null controllability regions, then the control law generated by algorithm H_{MP} is also close to the true optimal control.

3.5. *Application to the realistic example.* Figure 2 shows a comparison between the trajectories for the optimal control law and algorithm H_{MP} for Example 2.3. In this particular case, the optimal control law was computed off-line by solving a sequence of linear programming problems, while algorithm H_{MP} was limited to computation time compatible with an on-line implementation. The value of ϵ was set to 0.01 (using linear programming it was determined that the maximum value of ϵ compatible with the constraints is 0.025) and each of the regions C_{sk} was found as the convex hull of 32 points, using optimal trajectories generated off-line. Note that in spite of being limited to running time roughly two orders of magnitude smaller than the computation time used off-line to find the true optimal control solution, algorithm H_{MP} generates a solution that takes only 25% more time to get to the origin (25 vs 31 stages).

Figure 3 shows the results of applying algorithm H_{MP} when the heuristic is perfect (i.e. the exact transit time to the origin is known). By comparing Figs 2 and 3 we see that most of the additional cost comes from the approximation made in Theorem 2, while the use of an estimate of the cost-to-go based upon the subestimates C_{sk} (rather than a "true" underestimate as required by Theorem 1) adds only one stage to the total transit time.

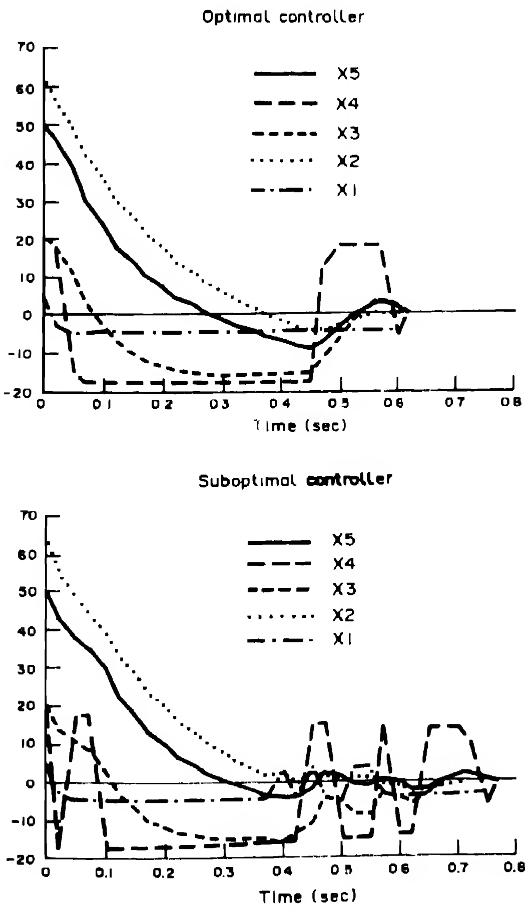


FIG. 2. Optimal control vs algorithm H_{MP} for the Example of Section 3.4.

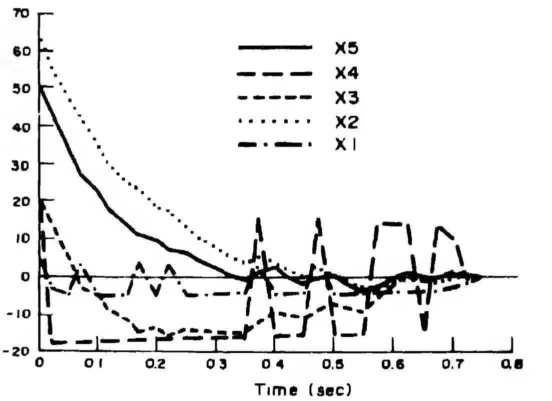


FIG. 3. Algorithm H_{MP} with perfect information.

4. *Conclusions*

Following the idea presented by Sznaier (1989) and Sznaier and Damborg (1990), in this paper we propose to address time-domain constraints by using a feedback controller based upon the on-line use of a dynamic-programming approach to solve a constrained optimization problem. Theoretical results are presented showing that this controller yields asymptotically stable systems, provided that the solution to an optimization problem, considerably simpler than the original, can be computed in real-time. Dimensionality problems common to dynamic programming approaches are circumvented by applying a suitably modified discrete time minimum principle, which allows for checking only the vertices of a polytope in control space. This polytope is obtained by considering the intersection Ω_1 of the original control region Ω with the region obtained by projecting the state constraints into the control space. The proposed approach results in a substantial reduction of the dimensionality of the problem (two orders of magnitude for the case of the example presented in Section 2.3). Hence, the proposed algorithm presents a significant advantage over previous approaches that use the same idea, especially for cases, such as Example 2.3, where the time available for computations is very limited.

We believe that the algorithm presented in this paper shows great promise, especially for cases where the dimension of the system is not small. Note however, that the algorithm requires the real time solution of two non-trivial computational geometry problems in R^n; determining the inclusion of a point in a convex hull and finding all the vertices of a polytope. Recent work on trainable non-linear classifiers such as artificial neural nets and decision trees may prove valuable in solving the first problem.

Perhaps the most serious limitation to the theory in its present form arises from the implicit assumptions that the model of the system is perfectly known. Since most realistic problems involve some degree of uncertainty, clearly this assumption limits the domain of application of the proposed controller. We are currently working on a technique, patterned along the lines of the Norm Based Robust Control framework introduced by Sznaier (1990), to guarantee robustness margins for the resulting closed-loop system. A future paper is planned to report these results.

References

Benzaouia, A. and C. Burgat (1988). Regulator problems for linear-discrete time systems with non-symmetrical constrained control, *Int. J. Control*, **48**, 2441–2451.

Bitsoris, G. and M. Vassilaki (1990). The linear constrained control problem. *Proc. 11th IFAC World Congress*, Tallin, Estonia, pp. 287–292.

Blanchini, F. (1990). Feedback control for linear time-invariant systems with state and control bounds in the

presence of disturbances. *IEEE Trans. Aut. Control.*, **35**, 1231–1234.

Boyd, S., V. Balakrishnan, C. Barrat, N. Khraishi, X. Li, D. Meyer and S. Norman (1988). A new CAD method and associated architectures for linear controllers. *IEEE Trans. Aut. Control*, **33**, 268–283.

Butkovskii, A. G. (1963). The necessary and sufficient conditions for optimality of discrete control systems *Automation and Remote Control*, **24**, 963–970.

Conway, J. B. (1990). *A Course in Functional Analysis*, Vol. 96 in Graduate texts in Mathematics. Springer-Verlag, New York.

Dreyfus, S. E. (1964). Some types of optimal control of stochastic systems. *J. SIAM Control A*, **2**, 120–134.

Fegley, K. A., S. Blum, J. O. Bergholm, A J. Calise, J. E. Marowitz, G. Porcelli and L. P. Sinha (1971). Stochastic and deterministic design via linear and quadratic programming. *IEEE Trans. Aut. Control*, **16**, 759–765.

Frankena, J. F. and R. Sivan (1979). A nonlinear optimal control law for linear systems. *Int. J. Control*, **30**, 159–178.

Gutman, P. O. and P. Hagander (1985). A new design of constrained controllers for linear systems. *IEEE Trans. Aut. Control*, **30**, 22–33

Gutman, P. O. and M. Cwikel (1986) Admissible sets and feedback control for discrete time linear dynamic systems with bounded controls and states. *IEEE Trans. Aut. Control*, **31**, 373–376.

Judd, R. P., R. P. Van Til and P. L Stuckman (1987) Discrete time quantized data controllers. *Int J. Control*, **46**, 1225–1233.

Sznaier, M. (1989). Suboptimal feedback control of constrained linear systems. Ph.D. Dissertation, University of Washington.

Sznaier, M. (1990). Norm based robust control of constrained discrete-time linear systems. *Proc. of the 29th IEEE CDC*, HI, 5–7 December, pp. 1925–1930.

Sznaier, M. and M. J. Damborg, (1990). Heuristically enhanced feedback control of constrained discrete-time linear systems. *Automatica*, **26**, 521–532.

Sznaier, M. (1991). Set induced norm based robust control techniques. In C. T. Leondes (Ed.), *Advances in Control and Dynamic Systems*, (to appear). Academic Press, New York.

Sznaier, M. and A. Sideris (1991a). Norm based optimally robust control of constrained discrete time linear systems. *Proc.* 1991 *ACC, Boston*, MA, pp. 2710–2715

Sznaier, M. and A. Sideris (1991b). Norm based robust dynamic feedback control of constrained systems. *Proc. of the First IFAC Symposium on Design Methods for Control Systems*, Zurich, Switerland, pp. 258–263.

Vassilaki, M., J. C. Hennet and G. Bitsoris (1988). Feedback control of discrete-time systems under state and control constraints. *Int J. Control*, **47**, 1727–1735.

Winston, P. H. (1984) *Artificial Intelligence*, 2nd ed Addison-Wesley, MA.

Zadeh, L A. and B H Whalen (1962) On optimal control and linear programming *IRE Trans Aut. Control*, **7**, 45–46.

Automatica, Vol. 29, No. 2, pp. 445–449, 1993
Printed in Great Britain.

0005–1098/93 $6.00 + 0.00

Brief Paper

Utility of Imaging Sensors in Tracking Systems*

D. D. SWORDER,† R. G. HUTCHINS‡ and M. KENT†

Key Words—Adaptive systems; gain control; image processing Kalman filters; target tracking.

Abstract—To justify the introduction of new sensor hardware, the potential performance improvement must be sufficient to warrant the additional expense and potential design risks that a change in system architecture entails. This paper compares the response of an image-enhanced tracker with that obtained from a conventional sensor suite. Because the underlying motion model is nonlinear and the tracking environment non-Gaussian, performance is scenario dependent. For a problem of tracking an agile target in the plane, it is shown that a system containing an imager is superior to an orthodox tracker without one.

1. *Introduction*

IT IS OFTEN DIFFICULT to predict the ultimate impact of hardware innovations on system design. When a novel device accomplishes a previously feasible task, its role is apparent. However, when it improves on one aspect of a task already achievable to some degree using standard methods, its utility is not as clear. To justify introducing new hardware, it must be evident that the ensuing system has capabilities significantly greater than those found in more familiar systems, and to achieve this improvement new architectures may be required. A situation of this type has developed as FLIR sensors have been introduced into tracking loops. An imaging sensor produces a sequence of "pictures" of the target. An image processor interprets these scenes, and identifies target attributes that are naturally manifested in properties of target extent. Figure 1 shows a tracking architecture proposed in several references as a means for integrating the information derived from an imager with that from conventional sensors (Kendrick *et al.*, 1981). The lower path shows the usual mapping from center-of-reflection observations to an estimate of target location. The upper path is image-specific, and uses observations of target shape to augment the point-location data in the lower path. Note that the information flow in the upper path is a complement to, not a duplicate of, that in the lower.

Merging these disparate data requires considerable care if the potential performance improvement is to be realized. The additional complexity and expense implicit in the architecture shown in Fig. 1 must be justified if imagers are to be accepted into tracking systems. This is not an easy task because the commonly used performance indices are difficult to compute for realistic representations of an encounter. The models required to faithfully portray the peculiarities of typical motion paths are nonlinear, non-Gaussian, and quite scenario dependent. To gain insight into the idiosyncracies of image-enhanced trackers, and to contrast them with those of their more established brethren, this paper considers the response in a specific vignette; an agile target moving in the plane with intermittent maneuvers is tracked from a fixed location. The object of this study is the quantification of the utility of the upper path in Fig. 1. The detailed conclusions are encounter-specific, but the substance of the results holds in more general circumstances.

2. *Model-based trackers*

In model-based trackers, as the name implies, an analytical encounter model is used as an intermediary in deriving algorithms for estimation and prediction. Analysts usually make no pretence regarding the completeness of the model; judging it on the basis of the adequacy of the algorithms to which it leads. In its simplest form, the model relates spatio-temporal measurements of target location through an ostensible motion equation which has a sufficiently intuitive parametrization to permit natural modification as the conditions of the encounter change. Model-based state estimation algorithms have a long history, and considerable experience exists which points to both the strengths and the weaknesses of the approach (Maybeck, 1982). The most common formulation uses a linear Gauss–Markov (LGM) motion model for the state process $\{x_t\}$. The observation process $\{y_t\}$ is often derived from a radar or similar device that provides a measurement—albeit a noisy one—of the center-of-reflection location of the target; e.g. range and bearing.

There is a well known solution to the resulting estimation problem—the (extended) Kalman filter. Denote the filtration generated by $\{y_t\}$ by $\{Y_t\}$. Then the conditional mean of the encounter state ($\hat{x}_t = E\{x_t \mid Y_t\}$) is given by the algorithm

$$\mathrm{d}\hat{x}_t = A\hat{x}_t\,\mathrm{d}t + P_{xx}D'R_v^{-1}\,\mathrm{d}\nu_t, \tag{2.1}$$

where A is the dynamic matrix in the motion model, D is the observation gain, R_v is the intensity of the observation noise, $\mathrm{d}\nu_t$ is the increment of the innovations process, and P_{xx} is the error covariance matrix. This algorithm has been used in non-LGM applications as well. Unfortunately, it is difficult to compare estimation algorithms in a non-Gaussian environment. The error covariance cannot be computed strictly, and even if a pseudo-covariance is computed in its stead, it is not always clear how well it quantifies estimator accuracy. To illustrate these ideas in the context of a specific example, consider the problem of tracking the motion of an agile target in the x–y plane. A simple motion model for an evasive target moving at essentially constant speed is:

$$\mathrm{d}\begin{pmatrix} X \\ Y \\ V_x \\ V_y \end{pmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\omega_\psi \\ 0 & 0 & \omega_\psi & 0 \end{bmatrix}\mathrm{d}t + \mathrm{d}\begin{pmatrix} w_x \\ w_y \end{pmatrix}, \tag{2.2}$$

where $\{X_t, Y_t\}$ are position coordinates, and $\{V_x, V_y\}$ are associated velocities. The target is assumed to be subject to two types of acceleration; a wide band omni-directional acceleration described by the Brownian motion $\{w_x, w_y\}$ of intensity W, and a maneuver acceleration represented by the turn rate process $\{\omega_\psi\}$. The former fits well within the LGM framework, but the latter is troublesome in two respects. First, it enters into the target dynamics as a state multiplier,

* Received 30 July 1991; revised 3 March 1992; received in final form 10 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor R. Boel under the direction of Editor H. Kwakernaak.

† Dept. of ECE, Univ. of California, San Diego, La Jolla, CA 92093-0407, U.S.A.
‡ ORINCON Corp, San Diego, CA 92121, U.S.A.

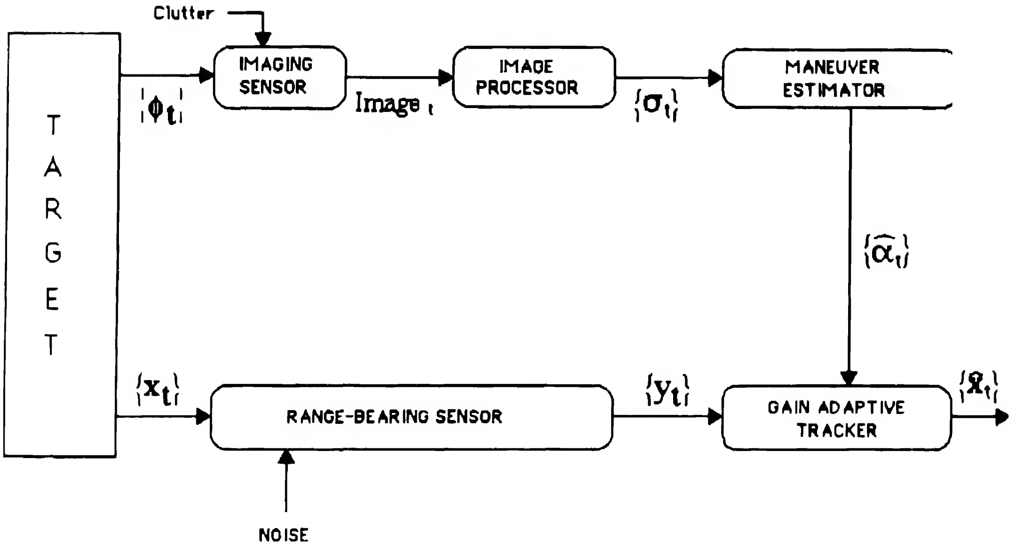


FIG. 1 Image enhanced estimation architecture

and consequently renders the dynamic equation nonlinear in the augmented state space Secondly the sample paths of the maneuver accelerations are usually described as jump processes, and are clearly non Gaussian (Cloutier *et al* 1989) In volatile encounters this lateral component is of primary importance thus making the LKF formalism of doubtful applicability (Speyer *et al* 1990)

To complete the motion model the dynamics of the turn rate process must be quantified Suppose that the set of possible maneuvers is partitioned into K levels $\omega_\psi \in \{a_1 \; a_K\}$ and let $\{\alpha_t\}$ be the maneuver indicator process, $\alpha_t = e_i$ if $\omega_\psi = a_i$ where e_i is the unit vector in the ith direction in the appropriate space The simplest way to delineate the maneuver process is to suppose the successive maneuver modes are represented by a Markov chain with transition rate matrix q $\mathcal{P}(\alpha_{t+\epsilon} = e_j \mid \alpha_t = e_i) = \delta_{ij} + q_{ij}\epsilon + o(\epsilon)$ This is a tractable model the parameters of which can be expressed in terms of the mean lifetime of each maneuver mode and the transition probabilities between modes

Before analysing a tracker with the full sensor suite consider an architecture which has no image path i e the tracker avails itself of just the range bearing measurement as shown in the lower path in Fig 1 These measurements will be taken at a fixed rate (10 samples $\sec^{-1}$) with independent errors Both the target model and the observation equation are nonlinear thus precluding the explicit evaluation of the conditional mean of the target state Additionally the maneuver accelerations have a non Gaussian jump character Nevertheless the need to achieve high accuracy tracking in applications like this had led to the development and implementation of plausible algorithms The most common approach is also the simplest, the maneuvers are neglected The EKF paradigm (linear motion nonlinear observation) is now directly applicable A four dimensional filter is sufficient to generate the state estimate (position and velocity) The filter gain depends parametrically on the intensity of the motion disturbance and the family of filters will be denoted by $\mathrm{EKF}_W(W)$ The performance of the estimator tends to be good if conditions are benign but by overlooking what is in fact the main ingredient in the acceleration, the filter is slow to recognize a turn-induced path change, and large errors build The basic reason for this belated response to lateral acceleration is that the motion equation neglects maneuvers thus leading to a small error covariance P_{tt} The simplest way to avoid the large error after maneuver onset is to arbitrarily increase W in equation (2 2), i e add pseudonoise (Maybeck, 1982) Unfortunately the same higher gain that makes the filter more responsive, and hence a better tracker during the maneuver is also a gain that magnifies the observation noise during quiescent conditions The exigencies created by the need for high accuracy tracking during quiescent operation must be balanced against the possibility of large errors and loss of lock during turns and with the single degree of freedom found in LKF_W this is difficult to do

A slight increase in model complexity gives a more realistic motion equation Rather than neglecting the turn rate it can be replaced by a Gaussian surrogate If the initial probability distribution is selected correctly $\{\omega_\psi\}$ can be represented by a wide sense stationary process with a power spectral density $\Phi_\psi(\omega)$ that can be computed using familiar techniques The resulting acceleration dynamics can be integrated into the position–velocity dynamics by increasing the size of both the state vector and the Brownian motion (Berg 1983) If the shaping filter is first order the turn rate model is a two parameter family with τ the time constant of the shaping filter and $R(0)$ $E\{(\omega_\psi)\}$ the intensity of the turn rate process This leads naturally to a two parameter family of filters $\mathrm{FKF}_s(R(0)\ \tau)$ with the correct choice depending on the volatility of the encounter Though the motion equation is nonlinear but it can be linearized using conventional methods Furthermore the acceleration dynamics are included explicitly, and the correlations in the acceleration direction are maintained during flight (Hepner and Geering 1990)

In the dual path tracking architecture shown in Fig 1 the sensor suite contains an imager in addition to the conventional range bearing sensors Imagers are attractive because they generate data that bear on an aspect of target motion different from that usually processed in a tracking loop An estimation algorithm proposed in Sworder and Hutchins (1992) used the shape information produced by the imager for a faster indication of turning motion than that obtainable from location based residuals The imager is a pattern classifier and its errors are misclassifications Suppose that the imager processor collects data frames at a rate of λ frames $\sec^{-1}$ and classifies target orientation into L equally spaced orientation bins The output of the processor can be written as an L dimensional counting process $\{z_t\}$ the ith component of which is the number of times the target has been placed in bin i on the interval $[0\ t]$ This sequence of counts (or symbols) can be interpreted by a temporal processor to give the relative likelihoods of the various turn rate hypotheses This development is carried out in Sworder and Hutchins (1990) and can be summarized as follows Let the orientation indicator process be given by $\{\rho_t\}$, $\rho_t = e_i$ if the current target orientation is in the ith bin The quality of the imager is determined by the $L \times L$ discernability matrix $\mathbf{P} = [\mathbf{P}_{ij}]$ where $\mathbf{P}_{ij}$ is the probability that symbol i will be generated by the processor if j is the target orientation at time of image creation, i e $\mathbf{P}_{ij} = \mathcal{P}(\Delta z_i \neq 0 \mid \rho = e_j)$ The fidelity of image interpretation is a function of the imager

qualities, $\mathbf{P}$ and λ, and on the volatility of the encounter. This latter can be described by a comprehensive $KL \times KL$ dimensional transition rate matrix Q which depends on the generator of the acceleration chain and the generator of the possibly random angular motions for a given acceleration (Sworder and Hutchins, 1989). Denote the filtration generated by $\{z_t\}$ by $\{Z_t\}$. The conditional probabilities of the various turn rate hypotheses are given by $\hat{\alpha}_t = (\mathbf{I}_K \otimes \mathbf{1}_L')\hat{\phi}_t$, subject to $d\phi_t = Q'\phi_t\,dt + (\mathrm{diag}(\phi_t) - \hat{\phi}_t\hat{\phi}_t')(\mathbf{1}_K \otimes \mathbf{I}_L)\Lambda'\,\mathrm{diag}(\Lambda(\mathbf{1}_K \otimes \mathbf{I}_L)\phi_t)^{-1}dz_t$ where $\Lambda = \lambda\mathbf{P}$, $\mathbf{I}$ is the identity matrix, and $\mathbf{1}$ is a unity vector.

In the tracker architecture proposed in Hutchins and Sworder (1992), the turn rate terms were treated as additive disturbances, and their Z_t conditional means were added to the conventional nonmaneuvering EKF. This proved to be a useful approach, but had poor performance subsequent to the maneuver. This deficiency follows directly from the fact that a small value of W results in a slow decay of any errors created during the maneuver. But the process noise intensity W can be augmented proportionally to acceleration uncertainty $(\mathrm{Var}\,\omega_{ij} = \Sigma_i a_i^2\hat{\alpha}_i - (\Sigma_i a_i\hat{\alpha}_i)^2)$ transformed into the Cartesian coordinate system. The corresponding value of P_x is that with the increased W. Through this simple artifice, a maneuver adaptive tracker is created. Denote this image enhanced adaptive filter by EKF_I.

3. A comparison of the filters for a specific vignette

It is difficult to compare the performance of the various filters described in Section 2. There is an intuitive sense that as the model becomes more representative of the actual motion, the performance of a tracker based upon it should improve. Still, it is not clear that EKF_I is enough better to justify the hardware/software expense required to implement it. To illustrate the contrasts in performance, consider a specific vignette with a target initially located at $(X_0, Y_0) = (1.0, 6.4)$KM and initial velocity $(v_{x0}, v_{y0}) = (5.0, -13.4)$ m sec^{-1}. Moving at constant velocity for 30 sec, a 0.5 g turn is made for 6 sec, after which it returns to constant velocity motion. The omnidirectional accelerations will be assumed to be slight: $(dw_x)^2 = (dw_y)^2 = 0.01\,dt$ (m sec^{-2})^2. Figure 2 shows the target path (truth) without the wide band acceleration.

Tracking will be accomplished with a 10 samples sec^{-1} range-bearing sensor located at the origin with Gaussian errors of standard deviation 5.0 m and 0.25°, respectively, independent in time and type. This measurement specification suffices to determine the basic EKF_W. If an imager is available, it will be collocated and make measurements at the same rate, i.e. $\lambda = 10$ frames sec^{-1}, placing the target in angular bins of width 22.5° (number of angular bins $L = 16$). To complete the measurement model, the imager errors must be quantified. The error taxonomy includes errors of three types: uniform errors in which the output symbol is uniformly distributed without regard to the true orientation, 1.0% in toto. Nearest neighbor errors in which the ostensible orientation is placed in the neighboring angular bin, 1.8% in toto. Silhouette projection errors in which the image processor places the target orientation in the bin situated symmetrically with respect to a line perpendicular to the line-of-sight, 10%. This last error is particularly important when image processing is dependent upon edge location because the target silhouette is invariant to the indicated shift. If the error categories are assumed to be independent in time and kind, the discernibility matrix can be deduced directly.

To determine EKF_I and the more sophisticated image-based tracker EKF_P, the maneuver dynamics must be specified. It will be supposed that $\{a_1 = 20°\,\mathrm{sec}^{-1}$ turn right, $a_2 =$ no turn, $a_3 = -20°\,\mathrm{sec}^{-1}$ turn left$\}$ and that the chain $\{\alpha_t\}$ is symmetric about the null mode, i.e. the probabilistic character of a_1, a_1 and a_3 are identical. The elements of the q matrix are determined jointly by the mean sojourn times in each acceleration mode $\{\nu_i, i = 1, 2, 3; \nu_1 = \nu_3\}$ and the transition probabilities from a maneuver to the nonmaneuvering mode. Let $q_1 = \mathcal{P}(\alpha_{t^-} = e_3 \mid \alpha_t = e_1$ and $\Delta\alpha_t \neq 0)$. Then q_1 measures the fraction of times that a maneuver ends in a coasting motion, e.g. $q_1 = 0$ implies pure jinking motion and $q_1 = 1$ always interjects coasting between turns. Specifically consider encounters with the tempos shown in Table 1. Encounter 3 is the most closely matched to the vignette. As mentioned earlier, the power spectral density can be computed for each encounter, and the shaping filter determined using conventional methods. Note that $R(0) = 2a_1^2\mathcal{P}(\alpha_t = e_1)$ can be calculated from the stationary distribution of the turn rate chain. Table 2 gives the parameters of the shaping filters for each encounter, and from this, the family of associated set of EKF_S can be determined.

Performance is difficult to quantify analytically. The motion model and the observation are nonlinear. Further, more, if the vignette is partitioned into pre, post, and maneuver coincident segments, the relative performance of the trackers differs as a function of the portion of the path that is being reviewed. A good indication of performance can, however, be determined by simulation. Figure 2 shows the position response of selected estimators in the x–y plane. To display tracker biases, a 20 trial mean sample path is shown rather than a (noisier) single sample path. The EKF_W used the indicated Brownian intensity, while both EKF_S and EKF_I are tuned to tempo 3. All of the filters perform well in the premaneuver phase, but both EKF_W and EKF_S have

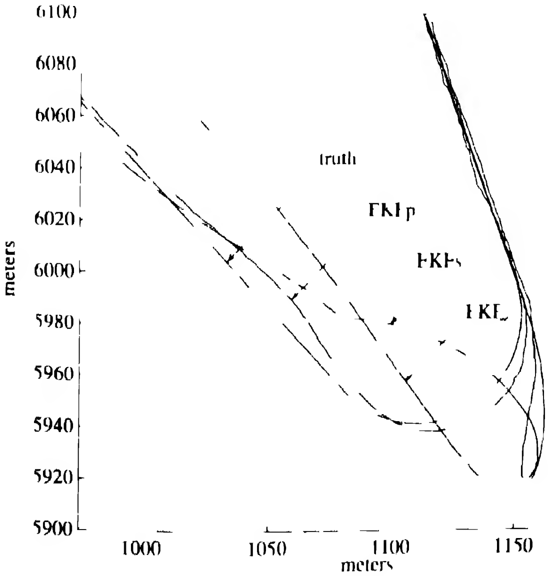


FIG. 2. Tracker performance in the plane.

TABLE 1. TEMPOS FOR THREE ENCOUNTERS

Encounter	Mean sojourn time	q_1	Comments
	$\nu_1 = \nu_3 = 4$ sec	0.0	Very volatile maneuver with no coasting
	$\nu_1 = \nu_2 = 4$ sec	0.5	Volatile maneuver with some coasting
	$\nu_1 = 4$ sec, $\nu_2 = 20$ sec	1.0	Less volatile with coasting between each maneuver

TABLE 2. PARAMETERS FOR THE SHAPING FILTER

Encounter	$R(0)$	Time constant, τ
1	$(20°\,\mathrm{sec}^{-1})^2$	2.0 sec
2	$(16.4°\,\mathrm{sec}^{-1})^2$	2.5 sec
3	$(5.76°\,\mathrm{sec}^{-1})^2$	6.25 sec

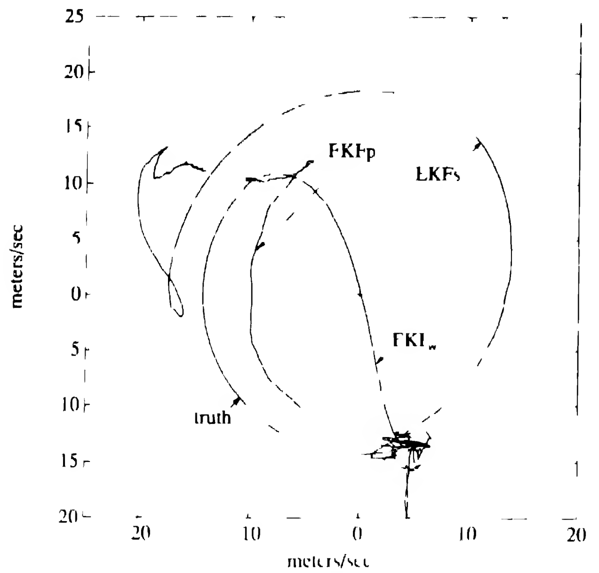


FIG. 3. Velocity estimates for different trackers.

difficulty following a turn, the latter making a pirouette at apogee. Only LKF_p performs satisfactorily in the post-maneuver initiation phase of the engagement. Figure 3 shows mean velocity profiles in the V_x-V_y plane and they are not as good. The filter tends to assign observation residuals to velocity to a much greater degree than one might expect. There is conspicuous misidentification of the velocity by both $EKF_W(0.01)$ and $EKF_s(5.76^2, 6.25)$ with neither recognizing the correct sense of rotation, the pirouette in the position path providing a velocity profile that is the mirror image of the actual path.

4. *Performance sensitivities*

The smoothed sample functions shown in the previous section give an indication of the relative performace of the various trackers. The estimators were selected on the basis of the actual encounter dynamics. Since both EKF_W and LKF_s are parametric families, it might be hoped that improved tracking could be obtained by tuning them to a pseudo-encounter, i.e. tuning for response rather than trying to mimic the encounter dynamics (Chang and Tabaczynski, 1989). If the vignette were not multimodal this would certainly be possible, but such *ad hoc* adjustment often achieves improved performance in one set of conditions at the expense of degraded performance elsewhere. To quantify performance, consider the paired tracking indices: the average error in the premaneuver phase, H_{av}, and the average peak error on the path, H_{max}. The former measures quiescent performance. When there are no maneuvers, it is obviously advantageous not to expect any, and so it would be expected that an EKF_W with a low gain (i.e. small W) would be attractive. The latter is the maximum error on the trajectory. From the sample curves of Section 3, it is evident that the time of occurrence of this maximum error will be during or soon after the maneuver. To minimize H_{max}, a high filter gain would be appropriate. These antithetical demands make tracker design difficult. The performance indices are difficult to evaluate analytically since the lack of LGM structure precludes finding the requisite moments. For this reason, the indices were determined using a simulation with a 50-trial sample average replacing the indicated expectation.

Figure 4 shows the performance of EKF_W for a range of W, along with the image-enhanced tracker EKF_p. In order to make the comparison of the filters easier, an icon has been used to display the performance of EKF_p. This filter is designed on the basis of a single intensity ($W = 0.01$), but the icon for H_{av} is shown to the left of the graph and that for H_{max} is shown to the right, thus placing it near the best performance attained with EKF_W as W varies from 0.01 to 100. Increasing W reduces the maximum position error from more than 100 m to less than 20 m. However quiescent error doubles for the same change. The best balance between nominal operation and maneuver tracking is achieved with a pseudonoise W around 10. Yet even with this tuning EKF_W is not as good as the image enhanced filter in either phase of the scenario. For example, for $EKF_W(10)$, $(H_{av}(0), H_{max}(0)) = (6, 13)$ m while for EKF_p the comparable figures are $(H_{av}(0), H_{max}(0)) = (4, 12)$ m. A 30% improvement in quiescent tracking is achieved while maximum tracking error is maintained at the same level.

Figure 5 shows four different performance curves for LKF_s with an iconic display for EKF_p as before. The intensity of the pseudonoise is related to the standard deviation of the turn rate and this is the abscissa of the curve. Again a trade off must be made since the time constant best suited to nominal operation is not so well adapted to sudden maneuvers. A good choice would be $(R(0), \tau) = (10, 2)$. With these parameters $(H_{av}(0), H_{max}(0)) = (5, 28)$ m which is comparable with the performance of $EKF_W(10)$ in the coasting phase but not nearly as good during a turn. It is interesting to note the strong resonance in Fig. 5 for $\tau = 20$ sec. Figure 6 shows that the maximum error is not maneuver coincident but occurs in the postmaneuver coast. The lag in the shaping filter causes the tracker to keep

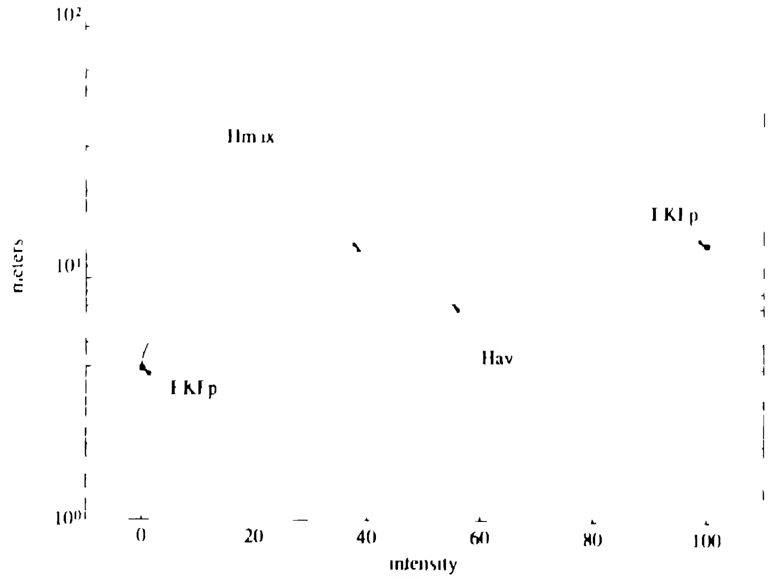


FIG. 4. Performance of EKF_W as a function of W.

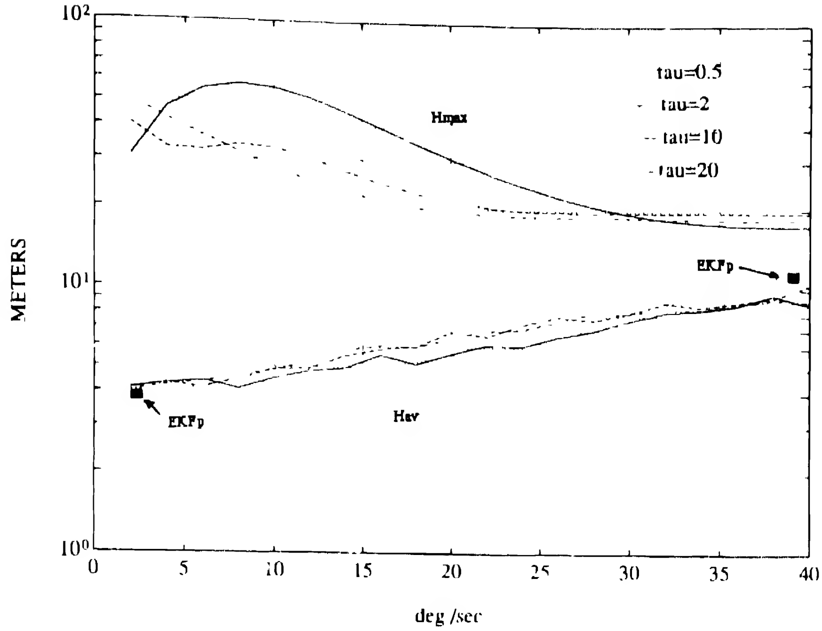


FIG. 5. Performance of EKF_s as a function of $(R(0))^{0.5}$ and τ.

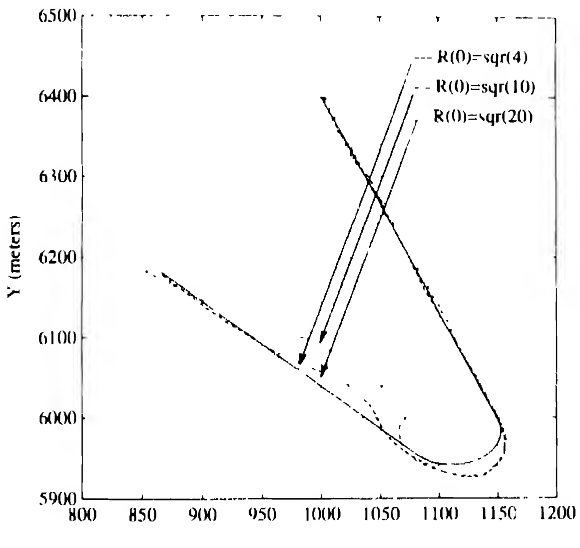


FIG. 6. Performance of $EKF_s(R(0), 20)$ for different values of $R(0)$.

turning after the maneuver has ceased, thereby creating a large overshoot.

5. *Conclusions*

New sensors require novel architectures to capture the potential for performance enhancement. It has been shown that the response of an image-enhanced tracking loop is superior to that attainable from conventional trackers without imaging capability. Because of the non-Gaussian nature of the problem, and the nonlinearities in the motion model, it is difficult to make inclusive generalizations about relative performance. It is evident though that an imager may achieve good quiescent performance without a severe turning penalty. This exchange is difficult to make with conventional algorithms. The two modified LGM procedures had similar performance, and in this application, increasing the order of the estimator to shape the ersatz Gaussian disturbance is not worth the additional complexity. This result is due primarily to the high accuracy of the location sensors; if the signal-to-noise ratio is high enough, even primitive models result in good estimators. The image-based architecture provides 40% better quiescent performance with a somewhat less pronounced improvement during a turn. This superior response certainly warrants capture

Acknowledgements—This research was supported by a grant from the Hughes Aircraft Co., and by the MICRO Program of the State of California under Project No. 91-156.

References

Berg, R. F. (1983). Estimation and prediction for maneuvering target trajectories. *IEEE Trans. Aut. Control*, **AC-28**, 294–304.

Chang, C. B. and J. A. Tabaczynski (1989). Application of state estimation to target tracking. *IEEE Trans. Aut. Control*, **AC-29**, 98–109.

Cloutier, J. R., J. H. Evers and J. J. Feeler (1989). Assessment of air-to-air missile guidance and control technology. *IEEE Control Systems Magazine*, **9**, 27–34.

Hepner, S. A. R. and H. P. Geering (1990). Observability analysis for target maneuver estimation via bearing-only and bearing-rate-only measurements. *J. of Guidance Control and Dynamics*, **13**, 977–983.

Hutchins, R. G. and D. D. Sworder (1992). Image fusion algorithms for tracking maneuvering targets. *J. of Guidance, Control and Dynamics*, **15**, 175–184.

Kendrick, D., P. S. Maybeck and J. G. Reid (1981). Estimation of aircraft target motion using orientation measurements. *IEEE Trans. Aerospace and Electronic Systems*, **AES-17**, 254–259.

Maybeck, P. S. (1982). *Stochastic Models, Estimation, and Control*. Vol. 2, Academic Press, New York.

Speyer, J. L., K. D. Kim and M. Tahk (1990). Passive homing missile guidance law based on new target maneuver models. *J. of Guidance, Control and Dynamics*, **13**, 803–812.

Sworder, D. D. and R. G. Hutchins (1989). Image-enhanced tracking. *IEEE Trans. Aerospace and Electronic Systems*, **AES-25**, 701–710.

Sworder, D. D. and R. G. Hutchins (1990). Maneuver estimation using measurements of orientation. *IEEE Trans. Aerospace and Electronic Systems* **AES-26**, 625–638.

Sworder, D. D. and R. G. Hutchins (1992). Improved tracking of an agile target. *J. of Guidance, Control and Dynamics*, **15**, 1281–1284.

Automatica, Vol. 29, No. 2, pp. 451–455, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Adaptive Frequency Response Identification Using the Lagrange Filter*

YU TANG†

Key Words—Identification; frequency response; filtering; adaptive control.

Abstract—This paper addresses the problem of identifying the frequency response of a stable discrete-time system at a set of arbitrary frequencies. The output of the identifier may be used directly in control synthesis based on points of frequency response, or to give a frequency response estimate together with an H_∞ bounding function for robust control. Central to our idea is Lagrange interpolation. Conditions for exponential convergence of the parameter estimates are given in terms of energy of input signal. Simulations are presented to illustrate the analytical results.

1. *Introduction*

RECENTLY, EFFORTS to combine robust control synthesis with adaptive techniques have been made to handle the robustness of adaptive control systems (e.g. Ortega and Tang, 1989, and the references cited there). A key element in such approaches is an estimator giving information on the frequency response of the plant, e.g. a set of points of frequency response, a frequency response estimate, and/or a bounding function of the modeling error. This paper deals with estimating frequency response at a set of arbitrary frequencies of interest.

Central to our idea is Lagrange interpolation. This gives a parametric model (interpolation model) which interpolates the frequency response at a set of user-defined arbitrary frequencies. Then the interpolation model is estimated using a recursive identification algorithm, and the frequency response estimate at these points is calculated via a linear map. These estimates may be used directly for on-line tuning of controllers (Åström and Hagglund, 1984; Tang, 1989) or, to be further elaborated, to give a frequency response estimate and/or a bounding function of the modeling error for robust control purposes (Kosut, 1988; Helmicki *et al.*, 1989; Gu and Khargonekar, 1992).

Frequency response identification via interpolation apparently was first done in Bitmead and Anderson (1981), where the frequency sampling filter, a special case of the Lagrange filter when the interpolation points are equally spaced around the unit circle, is used as the interpolation model. An H_∞ bound between the plant and the interpolation model was later established in Parker and Bitmead (1987a) and Helmicki *et al.* (1989). As pointed out in Parker and Bitmead (1987b) and Helmicki *et al.* (1989), in the presence of disturbances, the error between the estimated interpolation model and the plant is no longer uniformly bounded in N (the number of interpolation points). To remedy this, a class of two-stage identification schemes is proposed, see, e.g. Helmicki *et al.* (1989), Gu and Khargonekar (1992) and the references cited there.

* Received 19 June 1991; revised 25 June 1992; received in final form 1 July 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor R. R. Bitmead under the direction of Editor P. C. Parks.

† DEPFI-UNAM, P.O. Box 70-256, 04510 Mexico D.F., Mexico.

This paper extends the results of Bitmead and Anderson (1981) and Parker and Bitmead (1987a) to the case where the interpolation points are arbitrarily spaced around the unit circle. It is shown that if the input signal is rich in a frequency band, then the corresponding parameter estimates will converge to a neighborhood of their true values, which reduces to a single point (the true value) in the absence of the signal interference and external disturbances. The remainder of the paper is organized as follows: the Lagrange filter is described in Section 2. The identification scheme is given in Section 3. Section 4 is devoted to establish the convergence of the parameter estimates upon the input signal. Simulation results are presented in Section 5, and in Section 6, concluding remarks are given.

2. *The Lagrange filter*

Let $G(z)$ be a transfer function with real coefficients, analytic in $|z| > \rho_0$, $0 < \rho_0 < 1$, and $\{g(t)\}_{t=0}^{\infty}$ its impulse response. We want to approximate it by $G^I(z)$ by means of Lagrange interpolation (Geddes and Mason, 1975):

$$G^I(z) = \sum_{k=0}^{N-1} G(z_k) l_k(z), \tag{2.1}$$

where $\Omega := \{z_k = e^{jw_k}\}_{k=0}^{N-1}$, with $0 \le w_k < 2\pi$, is a set of N distinct points around the unit circle, and $l_k(z)$ are the interpolation filters,

$$l_k(z) = \frac{\prod_{m=0}^{N-1} (1 - z_m z^{-1})}{\prod_{m=0,\, m\neq k}^{N-1} (1 - z_m z_k^{-1})\,(1 - z_k z^{-1})}. \tag{2.2a}$$

It is easy to see that the frequency response of $l_k(z)$ satisfies

$$l_k(z_m) = \begin{cases} 1 & \text{for } m = k, \\ 0 & \text{for } m \neq k. \end{cases} \tag{2.2b}$$

Since the transfer function $G(z)$ has real coefficients, the information of the frequency response in $[0, \pi]$ is sufficient to determine the whole frequency response. We will consider the case where the interpolation points z_k are symmetrically spaced around the unit circle respect to the real axis with $z_0 = 1$. The case where $z_0 = e^{jw_0}$ for $w_0 \neq 0$ can be treated similarly.

Let Ω_s be the set of interpolation points symmetrically spaced around the unit circle respect to the real axis (N odd),

$$\Omega_s = \{z_k = e^{jw_k} \mid w_0 = 0,\ w_k = -w_{N-k},\ \text{for } k = 1, 2, \ldots, L = (N-1)/2\}. \tag{2.3}$$

Define

$$A_k = \mathrm{Re}[G(z_k)], \quad B_k = \mathrm{Im}[G(z_k)], \quad \text{for } k = 0, 1, \ldots, L, \tag{2.4}$$

$$\alpha_k = \frac{2^L}{N} \prod_{m=1}^{L} [1 - \cos(w_k)], \quad \beta_k = 0, \quad \text{for } k = 0, \tag{2.5a}$$

$$\alpha_k = \frac{2^l}{N}\cos\left(\frac{N}{2}w_k\right)[\cos(\tfrac{3}{2}w_k) - \cos(\tfrac{1}{2}w_k)]$$

$$\times \prod_{\substack{m=1 \\ m\neq k}}^{L} [\cos(w_k) - \cos(w_m)],$$

$$\beta_k = -\frac{2^l}{N}\sin\left(\frac{N}{2}w_k\right)[\cos(\tfrac{3}{2}w_k) - \cos(\tfrac{1}{2}w_k)]$$

$$\times \prod_{\substack{m=1 \\ m\neq k}}^{l} [\cos(w_k) - \cos(w_m)], \quad \text{for } k = 1, 2, \ldots, L, \tag{2.5b}$$

then the interpolation model (2.1) can be expressed as

$$G^l(z) = \sum_{k=0}^{l} (c_k + d_k z^{-1}) H_k(z), \tag{2.6}$$

where the coefficients are calculated from

$$c_k = A_k, \quad d_k = 0, \quad \text{for } k = 0, \tag{2.7a}$$

$$\begin{bmatrix} c_k \\ d_k \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -2\cos(w_k) & -2\sin(w_k) \end{bmatrix}\begin{bmatrix} \alpha_k & \beta_k \\ -\beta_k & \alpha_k \end{bmatrix}\begin{bmatrix} A_k \\ B_k \end{bmatrix}$$

$$\text{for } k = 1, 2, \ldots, L, \tag{2.7b}$$

$H_k(z)$ are bandpass filters with center frequency w_k,

$$H_k(z) = \frac{H_c(z)}{1 - z^{-1}} \quad \text{for } k = 0, \tag{2.8a}$$

$$H_k(z) = \frac{H_c(z)}{\alpha_k^2 + \beta_k^2}\,\frac{1}{1 - 2\cos(w_k)z^{-1} + z^{-2}}$$

$$\text{for } k = 1, 2, \ldots, L, \tag{2.8b}$$

and $H_c(z)$ is a comb filter,

$$H_c(z) = \frac{(1 - z^{-1}) \prod_{m=1}^{L} (1 - 2\cos(w_m)z^{-1} + z^{-2})}{N} \tag{2.9}$$

Therefore, given the frequency response at the interpolation points Ω_l (2.3), the interpolation model can be obtained from (2.6)–(2.9). Some important properties of the interpolation model are summarized in the following proposition, whose proof is given in Appendix A.

Proposition 2.1. Let $G(z)$ be analytic in $|z| > \rho_0$, $0 < \rho_0 < 1$, and $G^l(z)$ be the interpolation model given by (2.6)–(2.9). Then

(i) $G^l(z)$ is an $N-1$ degree polynomial in z^{-1}.

(ii) $G^l(z)$ interpolates $G(z)$ at $z_k = e^{jw_k}$ for $k = 0, 1, \ldots, N-1$, i.e.

$$G^l(z_k) = G(z_k), \quad k = 0, 1, 2, \ldots, N-1 \tag{2.10}$$

(iii) The approximation error is bounded by

$$\|G^l - G\|_\infty \le M\rho_0^N, \tag{2.11}$$

where $M = M(N, \rho, G) > 0$ grows linearly in N.

From (2.11) it follows that $\|G^l - G\|_\infty \to 0$, as $N \to \infty$. For their later use, we introduce the normalized bandpass filters (i.e. their frequency response at the center frequency is one) as

$$H_k'(z) = f_k(z)H_k(z), \quad k = 0, 1, \ldots, l, \tag{2.12a}$$

with $f_k(z)$ being the normalizing factors given by

$$f_k(z) = 1, \quad \text{for } k = 0, \tag{2.12b}$$

$$f_k(z) = 2\{\alpha_k - [\alpha_k \cos(w_k) - \beta_k \sin(w_k)]z^{-1}\},$$

$$\text{for } k = 1, \ldots, L \tag{2.12c}$$

It can be verified that

$$H_k'(z) = I_k(z), \quad \text{for } k = 0,$$

$$H_k'(z) = I_k(z) + I_{N-k}(z), \quad \text{for } k = 1, 2, \ldots, L,$$

with $I_k(z)$ the interpolation filter defined in (2.2).

Remark 2.1. A special case is when the interpolation points are equally spaced around the unit circle, i.e. $w_k = 2\pi k/N$, which gives a frequency sampling implementation of the interpolation model (Bitmead and Anderson, 1981). In this implementation of the interpolation model (Bitmead and Anderson, 1981). By simple calculations it can be shown that in (2.5), $\alpha_k = 1$, $\beta_k = 0$, for $k = 0, 1, \ldots, l$, and the comb filter (2.9) reduces to $H_c(z) = (1 - z^{-N})/N$.

Remark 2.2. Since there is a cancellation of pole-zero at the unit circle in (2.2a), to avoid instability of the filters, in the implementation of (2.8) and (2.12), z should be replaced by z/α_s, $0 < \alpha_s < 1$, which corresponds to a cancellation of pole–zero at the circle $|z| = \alpha_s$.

3. *Identification scheme*

3.1. *Plant assumptions and purposes.* Consider a linear time-invariant (LTI) discrete-time plant

$$y(t) = G(q)u(t) + d(t), \tag{3.1}$$

where $y(t)$, $u(t)$ and $d(t)$ are plant output, input and disturbances reflected at the output, respectively. $G(q)$ is the pulse transfer function of the plant. Assume that the plant is exponentially stable, i.e. the impulse response $g(t)$, $t > 0$, satisfies

$$|g(t)| \le M_g \rho_0^t, \quad \text{for } t > 0 \tag{3.2}$$

for some $M_g > 0$, and $0 < \rho_0 < 1$. This condition implies that the transfer function $G(z)$ is analytic in $|z| > \rho_0$ with no repeated poles at $|z| = \rho_0$ and $\sup_{|z|=\rho} |G(z)| \le M_g\rho/(\rho - \rho_0)$, for some $\rho \in (\rho_0, 1)$.

The purpose is to estimate the frequency response at $l + 1$ (l integer) arbitrary points $w_k \in [0, \pi)$ with starting point $w_0 = 0$. To this end, define

$$G(e^{jw_k}) = A_k + jB_k, \quad k = 0, 1, \ldots, l$$

The identification problem is then to estimate the parameters A_k and B_k from the input/output measurements. The remaining works are, (1) to design an identification scheme (i.e. an identifier structure and an estimation algorithm) (2) to choose an input with desired properties to achieve parameter convergence.

3.2. *Identifier*

Identifier structure

We shall consider the identifier implemented by the Lagrange filter. For this purpose, postulate an interpolation model (see equations (2.6)–(2.9))

$$G^l(q) = \sum (c_k + d_k q^{-1}) H_k(q) \tag{3.3}$$

It follows from (2.7) that the map $\{c_k, d_k\} \to \{A_k, B_k\}$ is given by

$$A_k = c_k, \quad B_k = 0, \quad \text{for } k = 0, \tag{3.4a}$$

$$\begin{bmatrix} A_k \\ B_k \end{bmatrix} = \frac{1/2}{\alpha_k^2 + \beta_k^2}\begin{bmatrix} \alpha_k + \beta_k \operatorname{ctg}(w_k) & \beta_k \csc(w_k) \\ \beta_k - \alpha_k \operatorname{ctg}(w_k) & -\alpha_k \csc(w_k) \end{bmatrix}\begin{bmatrix} c_k \\ d_k \end{bmatrix},$$

$$\text{for } k = 1, 2, \ldots, L. \tag{3.4b}$$

Therefore, the problem of identifying the $l + 1$ points of the frequency response boils down to estimating the parameters c_k and d_k of the interpolation model (3.3).

Estimation algorithm

We shall consider the normalized gradient algorithm to estimate the parameters c_k and d_k. Let

$$\theta_k = [c_k \;\; d_k]^T,$$

be the parameter vector for the band identifier k, and $\hat\theta_k(t)$ its estimate at time instant t, which is updated by

$$\hat\theta_k(t+1) = \hat\theta_k(t) + \gamma_k \frac{\phi_k(t)\varepsilon_k(t)}{1 + \phi_k^T(t)\phi_k(t)}, \quad k = 0, 1, \ldots, L, \tag{3.5}$$

where

$$\phi_k(t) = [u_k(t) \;\; u_k(t-1)]^T, \tag{3.6}$$

$$u_k(t) = H_k(q)u(t), \tag{3.7}$$

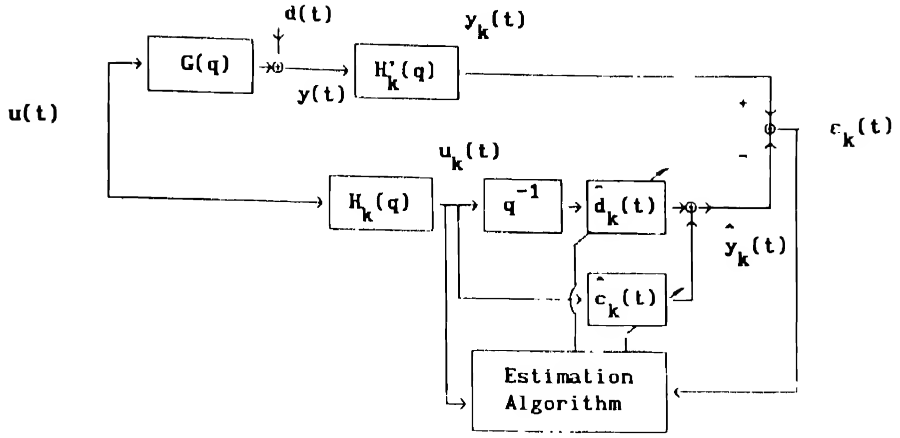


FIG. 1. A two-dimensional band identifier.

is the regressor,

$$\varepsilon_k(t) := y_k(t) - \hat{y}_k(t), \tag{3.8a}$$

with

$$y_k(t) = H'_k(q)y(t), \tag{3.8b}$$

$$\hat{y}_k(t) = \varphi_k^T(t)\hat{\theta}_k(t), \tag{3.8c}$$

is the estimation error, where $H'_k(q)$ are the normalized bandpass filters defined in (2.12) γ_k is a design parameter, $0 < \gamma_k < 2$. The identification scheme is depicted in Fig. 1.

Remark 3.1. The use of the normalized bandpass filters in (3.8b) is for obtaining the mainlobe error of the frequency response of $H'_k(q) - 1$ to be zero at w_k (see (B.3) of Appendix B).

Remark 3.2. In some cases, the impulse response may be of interest. Let $\{\hat{g}(t)\}_{t=0}^{N-1}$ be the impulse response of $G'(z)$, which can be calculated, given the frequency response of $G(z)$ at N points w_k, as follows: from (2.10) we have

$$G(e^{jw_k}) = \sum_{t=0}^{N-1} \hat{g}(t)e^{-jw_k t}. \tag{3.9}$$

Multiplying $e^{jw_k m}$ in both sides of (3.9) and summing it in k from 0 to $N-1$, gives

$$\kappa_m := \frac{1}{N}\sum_{k=0}^{N-1} G(e^{jw_k})e^{jw_k m} = \sum_{t=0}^{N-1} h_{mt}\hat{g}(t), \tag{3.10}$$

where

$$h_{mt} := \frac{1}{N}\sum_{k=0}^{N-1} e^{jw_k(m-t)} = \frac{1}{N}\left[1 + 2\sum_{k=1}^{l}\cos(m-t)w_k\right],$$

$$m\ t = 0, 1, \ldots, N-1. \tag{3.11}$$

Let $[h_{mt}]$ denote the matrix whose elements are h_{mt}, similarly, $[\kappa_m]$, $[\hat{g}(t)]$ the vectors whose elements are κ_m, $\hat{g}(t)$, respectively. Note that $[h_{mt}]$ is doubly symmetric (Toeplitz, and symmetric), with $h_{mm} = 1$, and $|h_{mt}| < 1$, for $m \neq t$, and is invertible if the interpolation points are distinct. Therefore, $[\hat{g}(t)]$ can be obtained from (3.10) by

$$[\hat{g}(t)] = [h_{mt}]^{-1}[\kappa_m]. \tag{3.12}$$

For the case where the interpolation points are equally spaced around the unit circle, $\{\hat{g}(t)\}_{t=0}^{N-1}$ and $\{G(e^{jw_k})\}_{k=0}^{N-1}$ form a pair of Discrete Fourier Series, and the calculation of (3.12) may be carried out using Fast Fourier transform.

4. *Analysis*

The purpose of this section is to establish conditions on the input signal to achieve convergence of the parameter estimates. First, we define the richness of a signal in terms of its energy in a frequency band.

Definition 4.1. Let $x(t)$ be a stationary signal with energy spectrum $S_x(e^{jw})$. It is said that $x(t)$ is rich in the frequency band k, centered at w_k and bandwidth $2\bar{w}_k$, if for some positive constant σ_k

$$S_x(e^{jw})\,dw \geq \sigma_k.$$

Let the parameter error vector $\tilde{\theta}_k(t)$ be defined as

$$\tilde{\theta}_k(t) := \hat{\theta}_k(t) - \theta_k = [\tilde{c}_k(t)\quad \tilde{d}_k(t)]' \tag{4.1}$$

The following theorem states the main results, whose proof is in Appendix B.

Theorem 4.1 Consider the identification scheme (3.3)–(3.8). Suppose that the plant (3.1) satisfies (3.2).

(i) If $u(t)$ is rich in the frequency band k, then $\hat{\theta}_k(t) \to \Theta_k := \{\theta \in \mathbb{R}^2 \mid |\theta| \leq R_k\}$ exponentially, with $R_k = R_k(N, \gamma_k, \bar{d}_k)$ a positive constant, and $\bar{d}_k$ the bound on the disturbances $d_k(t)$.

(ii) If $u(t)$ is a sine wave with frequency $w = w_k$ and $d_k(t) = 0$, then $\tilde{\theta}_k(t) \to 0$ exponentially.

Remark 4.1. Notice that when the input signal is rich in the frequency band k, the parameter estimate $\hat{\theta}_k(t)$ converges

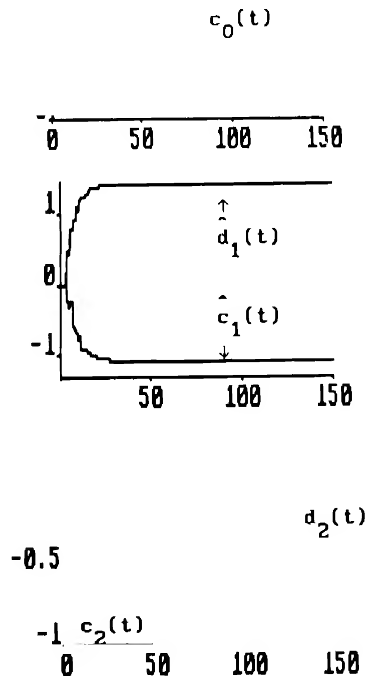


FIG. 2. Parameter estimates with $u(t)$ given in (5.2) (a) $r_0 = 2$, $r_1 = r_2 = 0$, (b) $r_1 = 2$, $r_0 = r_2 = 0$, (c) $r_2 = 2$, $r_0 = r_1 = 0$.

to a neighborhood of its true value θ_k with radius R_k, whose magnitude depends on the signal interference and external disturbances (see (B.9)). In the absence of the signal interference and disturbances, this neighborhood reduces to a single point (true value).

Remark 4.2. It is well known that when the input signal is not rich in a frequency band k, the parameter estimate in this band may diverge due to signal interferences and external disturbances. One solution to this is monitoring the richness of the input signal in each band, and tune the adaptation off in a frequency band when the signal to noise ratio is less then certain value (Tang and Fernandez, 1992).

5. *Simulation results*

The proposed identification scheme was used to identify the frequency response of the plant

$$G(q) = \frac{0.76q^2}{q^3 - 1.08q^2 + 0.264q \times 0.032}, \tag{5.1}$$

at $w_k = 2\pi k/5$, for $k = 0, 1, 2$. The plant (5.1) has poles at $\{0.8, 0.14 \pm j0.14\}$. The purpose of the simulations is to illustrate the main features of the identification scheme. To this end, the input signal was chosen as

$$u(t) = r_0 + r_1 \sin(w_1 t) + r_2 \sin(w_2 t), \tag{5.2}$$

and the parameters

$$N = 5, \quad \alpha_s = 0.99 \text{ (stability scalar, see Remark 2.2)},$$
$$\gamma_k = 0.5, \quad w_k = 2\pi k/5, \quad \text{for } k = 1, 2.$$

All initial conditions in the estimation law (3.5) were set zero. The true but unknown parameters are $c_0 = 5.00$; $c_1 = -1.06$, $d_1 = 1.41$; $c_2 = -0.69$, $d_2 = -0.48$. The behavior of the parameter estimates when $u(t)$ is a sine wave is shown in Fig. 2. Note that the parameter estimates converge to their true values as predicted by (ii) of Theorem 4.1. The evolution of the parameter estimates when the input is a pseudorandom binary signal (PRBS) is shown in Fig. 3. Due to the signal interference, the parameter estimates converge to a neighborhood of the true values, as predicted by (i) of Theorem 4.1.

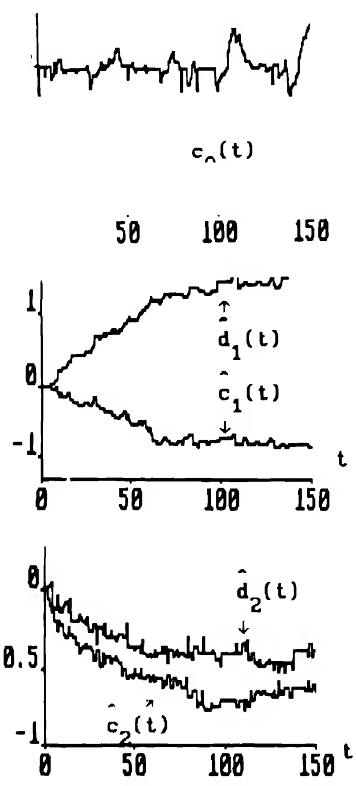


FIG. 3. Parameter estimates with $u(t)$ a PRBS.

6. *Conclusions*

An identification for identifying a set of $(L + 1)$ arbitrary frequences of an LTI discrete-time system has been proposed. The identifier consists of $L + 1$ band-identifiers, each works in its corresponding frequency band giving an estimate of the frequency response at the center frequency of this band. It was shown that if the input signal is rich in a frequency band, then the parameters of the corresponding band will converge to a neighborhood of their true values.

References

Anderson, B. D. O., R. R. Bitmead, C. R. Johnson, Jr., P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly and B. D. Riedle (1986). *Stability of Adaptive Systems: Passivity and Averaging Analysis*. MIT Press, MA.

Åström, K. J. and T. Hagglund (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, **20**, 645–651.

Bitmead, R. R. and B. D. O. Anderson (1981). Adaptive frequency sampling filters. *IEEE Trans. Acoustics, Speech, and Signal Processing*, **ASSP-29**, 684–693.

Geddes, K. O. and J. C. Mason (1975). Polynomial approximation by projections on the unit circle. *SIAM J. Numer. Anal.*, **12**, 111–121.

Gu, G. X. and P. P. Khargonekar (1992). A class of algorithms for identification in H_∞. *Automatica*, **28**, 299–312.

Helmicki, A. J., C. A. Jacobson and C. N. Nett (1989). H_∞ identification of stable systems: a scheme with direct application to control design. *American Control Conf. (ACC)*, Pittsburgh, PA, pp. 1428–1434.

Kosut, R. L. (1988). Adaptive control via parameter set estimation. *Int. J. Adaptive Contr. Signal Processing*, **2**, 371–399.

Ortega, R. and Y. Tang (1989). Robustness of adaptive controllers—a survey. *Automatica*, **25**, 651–677.

Parker, P. J. and R. R. Bitmead (1987a). Approximation of stable and unstable systems via frequency response interpolation. *Proc. 10th IFAC Congr. Automat. Contr.*, Munich, **10**, 358–363.

Parker, P. J. and R. R. Bitmead (1987b). Adaptive frequency response identification. *Proc. 26th Conf. Decision Control (CDC)*, Los Angeles, CA, pp. 348–253.

Tang, Y. (1989). A frequency domain approach to robust adaptive control. *Proc. American Control Conf. (ACC)*, Pittsburgh, PA, pp. 2285–2290.

Tang, Y. and E. Fernandez (1992). Adaptive pole placement: a frequency domain approach. *Proc. American Control Conf. (ACC)*, Chicago, IL, pp. 1466–1470.

Appendix A. Proof of Proposition 2.1

Proof of Proposition 2.1. The first two points follow straight from the Lagrange interpolation formula (2.1) and (2.2). The proof of (iii) is similar to that of Geddes and Mason (1975), and is given in the following.

Let

$$\mathcal{A}(\rho_0) := \{f : \mathbb{C} \to \mathbb{C} \mid f \text{ is analytic in } |z| > \rho_0,\ 0 < \rho_0 < 1\}, \tag{A1}$$

$$\mathcal{P}_N := \{\text{polynomials of order} \le N - 1\}, \tag{A2}$$

and $\mathcal{L}_N : \mathcal{A}(\rho_0) \to \mathcal{P}_N$ denote the Lagrange interpolation operation,

$$G'(z) = \mathcal{L}_N(G). \tag{A3}$$

According to (A3), and (3.9)–(3.12), the induced norm of $\mathcal{L}_N$ is bounded by

$$\begin{aligned} \|\mathcal{L}_N\|_\infty &:= \sup_{\|G\|_\infty} \|\mathcal{L}_N(G)\|_\infty = \sup_{\|G\|_\infty = 1} \|G'\|_\infty \\ &= \sup_{\|G\|_\infty = 1} \sup_{|z| = 1} |G'(z)| \\ &\le \sup_{\|G\|_\infty = 1} \sum_{i=0}^{N-1} |g_i| \le N \, \|[h_{mi}]^{-1}\|_1, \end{aligned} \tag{A4}$$

where $\|\cdot\|_1$ denotes the matrix one-norm. Therefore,

$$\|G - G'\|_\infty \le (1 + \|\mathcal{L}_N\|_\infty)\,\mathrm{dist}\,(G, \mathcal{P}_N) \le [1 + N\,\|[h_{mt}]^{-1}\|_1]M_1\rho_0^N, \quad \text{(A5)}$$

with

$$M_1 = M_g/(1 - \rho_0), \quad \text{(A6)}$$

and ρ_0, M_g are given in (3.2). The last inequality in (A5) follows from

$$\mathrm{dist}\,(G, \mathcal{P}_N) := \min_{p \in P_N} \|G - p\|_\infty \le \sum_{t=N}^{\infty} |g(t)| \le M_g\rho_0^N(1 - \rho_0).$$

The proof is completed with $M := [1 + N\,\|[h_{mt}]^{-1}\|]M_1$.

Appendix B. Proof of Theorem 4.1

It follows from (3.8b), (3.1), (2.12) and (2.8) that

$$y_k(t) = \phi_k^T(t)\bar{\theta}_k(t) + \eta_k(t), \quad \text{(B1)}$$

where

$$\eta_k(t) = H_k^m(q)u_k(t) + d_k(t), \quad d_k(t) = H_k'(q)\,d(t), \quad \text{(B2)}$$

and $H_k^m(q)$ is given by

$$H_q^m(q) := f_k(q)[G(q) - G'(q)] + f_k(q) \sum_{\substack{n=0 \\ n \ne k}}^{\cdot} (c_k + d_k q^{-1}) \times H_m(q) + (c_k + d_k q^{-1})[H_k'(q) - 1]. \quad \text{(B3)}$$

Note that in (B3) the first term arises from the approximation error $G(q) - G'(q)$, the second term is due to sidelobe effect of the bandpass filter $H_m(q)$, $m \ne k$, in the frequency band k, and the third term is mainlobe error $H_k'(z) - 1$. The use of the normalized bandpass filters in (3.8b) is now clear. The frequency response of $H_k^m(q)$ satisfies the following lemma.

Lemma B.1. For a number of interpolation N given, the frequency response of $H_k^m(q)$ in (B3) satisfies

(i) $\|H_k^m\|_\infty \le M_m < \infty$, with $M_m = M_m(N) > 0$ a constant.

(ii) $|H_k^m(z_k)| = 0$, with z_k an interpolation point.

Proof. (i) It follows from (B3) that

$$\|H_k^m\|_\infty \le \|f_k\|_\infty \|G - G'\|_\infty + \|f_k\|_\infty \sum_{\substack{m=0 \\ m \ne k}}^{l} (|c_k| + |d_k|)\,\|H_m\|_\infty + (|c_k| + |d_k|)\,\|H_k' - 1\|_\infty := M_m(N),$$

because for an N given, all the quantities in the right-hand side are bounded.

(ii) It follows that the frequency response of $G(q) - G'(q)$, $H_m(q)$, for $m \ne k$, and $H_k'(q) - 1$ are zero at $z_k = e^{jw_k}$.

Proof of Theorem 4.1. From (3.5), (3.8) and (B1) it follows that

$$\bar{\theta}_k(t + 1) = \Gamma_k(t)\bar{\theta}_k(t) + v_k(t), \quad \text{(B4)}$$

where

$$\Gamma_k(t) := I - \gamma_k \frac{\phi_k(t)\phi_k^T(t)}{1 + \phi_k^T(t)\phi_k(t)}, \quad v_k(t) := \gamma_k \frac{\phi_k(t)\eta_k(t)}{1 + \phi_k^T(t)\phi_k(t)}. \quad \text{(B5)}$$

Resolving (B4) gives

$$\bar{\theta}_k(t) = \Phi_k(t, 0)\bar{\theta}_k(0) + \sum_{\tau=0}^{t-1} \Phi_k(t, \tau + 1)v_k(\tau), \quad \text{(B6)}$$

with $\Phi_k(t, \tau)$ the transition matrix of (B4),

$$\Phi_k(t, \tau) := \prod_{i=\tau}^{t-1} \Gamma_k(i). \quad \text{(B7)}$$

Since $u_k(t)$ is a narrow-band signal, it can be approximated by a sine wave with frequency w_k. Hence, $\phi_k(t)$ is persistently exciting of order 2, which insures the exponential stability of (B4) (Anderson *et al.*, 1986). So there exist $K_\Gamma > 0$, and $|\lambda_k| < 1$, such that

$$|\Phi_k(t, \tau)| \le K_\Gamma |\lambda_k|^{t-\tau}, \quad \forall t \ge \tau \ge 0, \quad \text{(B8)}$$

which together with (B6) gives

$$|\bar{\theta}_k(t)| \le K_\Gamma\,|\lambda_k|^t\,|\bar{\theta}_k(0)| + K_\Gamma\,\|v_k\| \sum_{\tau=0}^{t-1} |\lambda_k|^\tau \le K_\Gamma |\lambda_k|^t\,|\bar{\theta}_k(0)| + \gamma_k K_\Gamma M_k' \frac{1 - |\lambda_k|^t}{1 - |\lambda_k|}, \quad \text{(B9)}$$

for some constant $M_m' = M_m'(N, d_0) > 0$. This proves the first part of the theorem.

(ii) If $u(t)$ is a sine wave with frequency w_k and $d_k(t) = 0$, it follows from (B2) and Lemma B.1 that $\eta_k(t) = 0$. This implies that $M_m' = 0$ in (B9). Hence, the conclusion follows from (B9).

Automatica, Vol. 29, No. 2, pp. 457–463, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Brief Paper

A Parametrization Approach to Optimal H_2 and H_∞ Decentralized Control Problems*†

RANJIT A. DATE‡ and JOE H. CHOW§

Key Words—Decentralized control; robust control; optimization; numerical methods.

Abstract—The paper proposes a two-stage design approach to optimal H_2 and H_∞ decentralized control problems. In the first stage, an optimal centralized controller is computed. Then in the second stage, based on the optimization results for centralized controllers, the parameter that decentralizes the controller is optimized. The design approach is applied to an observer-based decentralized controller. Optimality conditions are derived and examples are given to illustrate the design.

1. *Introduction*

THIS PAPER PROPOSES a two-stage optimal H_2 and H_∞ decentralized controller design in which the first stage involves an optimal centralized design and the second stage involves the optimization of the parameters that decentralize the optimal centralized controller. The approach is motivated by the optimal H_2 and H_∞ controller parametrization results in Doyle *et al.* (1989). This is in contrast to most decentralized controller designs (Davison and Gesing, 1979; Davison and Fergusson, 1981; Geromel and Bernusson, 1982; Uskokovic and Medanic, 1985; Bernstein, 1987; Iftar and Ozguner, 1989; Veillette, 1990; Siljak, 1991) in which the gains of the controllers are optimized directly. In this paper, we select the concurrent decentralized observer controller proposed in Veillette (1990) and Date and Chow (1989, 1991) to demonstrate the two-stage design approach, and to show the simplifications for computing suboptimal controllers. However, the design can be extended to other decentralized parametrizations.

We consider H_2 and H_∞ decentralized control design for two-channel systems. The results here can readily be generalized to systems with more than two channels. Let the model of the "generalized plant" be

$$\begin{aligned} \dot{x} &= Ax + B_w w + B_1 u_1 + B_2 u_2, \\ z &= C_z x + D_{121} u_1 + D_{122} u_2, \\ y_1 &= C_1 x + D_{211} w, \\ y_2 &= C_2 x + D_{212} w, \end{aligned} \tag{1}$$

where $x \in \mathscr{R}^n$ is the state variable, $w \in \mathscr{R}^{m_w}$ represents the disturbance or other external signals, $z \in \mathscr{R}^{n_z}$ represent the error outputs, and $u_i \in \mathscr{R}^{m_i}$ and $y_i \in \mathscr{R}^{p_i}$ are the controlled input and measured output of channel i, $i = 1, 2$. In input–output form, system (1) can be represented as

$$\begin{bmatrix} z \\ y_1 \\ y_2 \end{bmatrix} = G(s) \begin{bmatrix} w \\ u_1 \\ u_2 \end{bmatrix} = G(s) \begin{bmatrix} w \\ u \end{bmatrix}, \tag{2}$$

where the transfer function in packed matrix form is

$$G(s) \leftrightarrow \left[\begin{array}{c|ccc} A & B_w & B_1 & B_2 \\ \hline C_z & 0 & D_{121} & D_{122} \\ C_1 & D_{211} & 0 & 0 \\ C_2 & D_{212} & 0 & 0 \end{array}\right] \leftrightarrow \left[\begin{array}{c|cc} A & B_w & B \\ \hline C_z & 0 & D_{12} \\ C & D_{21} & 0 \end{array}\right]. \tag{3}$$

The symbol "$\leftrightarrow$" denotes the state space realization of a transfer function.

Assumption 1. System (1) is assumed to be in the form of a "standard problem" (Doyle *et al.*, 1989) that is

(1) (A, B_w) is stabilizable and (A, C_z) is detectable,
(2) (A, B) is stabilizable and (A, C) is detectable,
(3) $D_{12}^T D_{12} = I$, $D_{21} D_{21}^T = I$,
(4) $D_{12}^T C_z = 0$, $D_{21} B_w = 0$.

A decentralized controller for system (1) has the structure

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} K_1(s) & 0 \\ 0 & K_2(s) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = K_d(s) y. \tag{4}$$

This following assumption more than Part 2 of Assumption 1 will guarantee the existence of a decentralized controller with the structure (4) for system (1) (Wang and Davison, 1973).

Assumption 2. System (1) has no unstable decentralized fixed eigenvalues under the control structure (4).

Denote the closed-loop transfer function of the system using a decentralized controller (4) by $T_{zw}(K_d(s)) = T_{zw}^d$, and the set of all decentralized stabilizing controllers by $\mathscr{K}_d$. The decentralized H_2 optimal control problem is

$$\min_{K_d(s) \in \mathscr{K}_d} \|T_{zw}^d\|_2, \tag{5}$$

and the decentralized H_∞ suboptimal control problem is to find $K_d(s) \in \mathscr{K}_d$ such that

$$\|T_{zw}^d\|_\infty < \gamma, \tag{6}$$

where γ is a pre-specified constant.

Since a centralized controller $u = K(s)y$ will be designed in the first stage of the two-stage design, we denote the centralized closed-loop transfer function by $T_{zw}(K(s)) = T_{zw}$.

The paper is organized as follows. In Section 2, a decentralized controller parametrization is stated and the parametrization of decentralized controllers having the concurrent observer structure is derived. Section 3 contains the proposed H_2 design and Section 4 the H_∞ design. Design examples are given in these sections.

* Received 22 November 1991; revised 25 July 1992; received in final form 30 July 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor M. Ikeda under the direction of Editor A. P. Sage.

† The research is supported in part by NSF Grant ECS-8915667.

‡ Precision Automation and Robotics Pvt. Ltd., Pune, India.

§ Electrical, Computer and System Engineering Department, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, U.S.A.

2 A decentralized controller parametrization

Several decentralized controller parametrizations based on stable factorization (Gündeş and Desoer, 1990, Manousiaouthakis, 1989, Ozgüler, 1990, Date and Chow, 1991) are now available In Gundeş and Desoer (1990) and Date and Chow (1991), the parametrization uses a decentralized controller as the central controller in the Youla parametrization (Francis, 1987) An optimal design would require a direct optimization on the central controller In Ozguler (1990) the parametrization is based on a sequential feedback control of the decentralized channels The sequential design would naturally lead to an iterative optimization of controllers

The parametrization result in Manousiouthakis (1989) allows a centralized controller as the central controller and requires a parameter to decentralize the resulting controller Utilizing the H_2 and H_∞ design results in Doyle *et al* (1989), a two stage design approach can be formulated In the first stage, an optimal (or suboptimal) centralized controller is obtained based on the formulas in Doyle *et al* (1989) In the second stage, the decentralizing parameter is optimized In H_2 optimal design, in contrast to the finite order of optimal centralized controllers optimal decentralized controllers for strongly connected systems (Siljak 1991) may be infinite order This is commonly known as the *second guessing* phenomenon (Sandell *et al* 1978) As a result, the decentralizing parameter is also infinite order In this paper, we will investigate a concurrent decentralized observer controller with a finite order decentralizing parameter

The structure of the concurrent observer controller (Date and Chow, 1989) is given by

$$K_i(s) \leftrightarrow \begin{bmatrix} \hat{A} - B\bar{F} - L_iC_i & L_i \\ -\bar{F}_i & 0 \end{bmatrix}, \quad i = 1, 2 \tag{7}$$

where

$$\bar{F} = \begin{bmatrix} \bar{F}_1 \\ \bar{F}_2 \end{bmatrix} \tag{8}$$

and $\hat{A}$ contains a model of A and F_i and L_i, $i = 1, 2$ have the interpretation of full state feedback gains and decentralized observer gains The controller (7) is said to be concurrent because the gains F_i and L_i, $i = 1, 2$ are designed simultaneously and not sequentially Note that Assumption 2 is only a necessary condition for the existence of stabilizing concurrent observer controller The structure of $\hat{A}$ for the H_2 and H_∞ design problems will be specified later

Consider the controller $K(s)$ parametrization shown in Fig 1, where

$$M(s) \leftrightarrow \left[\begin{array}{c|cc} \hat{A} - BF - LC & L & B \\ \hline -F & 0 & I \\ -C & I & 0 \end{array}\right] \tag{9}$$

$Q_d(s) \in \mathscr{H}$ the set of stable proper transfer function matrices of appropriate dimensions and

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \quad L = [L_1 \quad L_2] \tag{10}$$

For $Q_d(s) = 0$ we obtain the centralized central controller

$$K(s) \leftrightarrow \begin{bmatrix} \hat{A} - BF - LC & L \\ -F & 0 \end{bmatrix} \tag{11}$$

In the two stage decentralized control design, the centralized gains F and L are optimized over all stabilizing gains

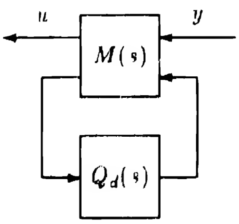


FIG 1 Controller $K(s)$ parametrization

We denote by

$$\mathscr{Q}_d = \{Q_d(s) \in \mathscr{H} \mid K(s) = K_d(s) \in \mathscr{K}_d\}, \tag{12}$$

to be the set of all the stable parameters $Q_d(s)$ that decentralizes the central controller and achieves the concurrent observer controller structure (7) Although the construction of a decentralizing parameter $Q_d(s)$ requires, in general, the solution to a Riccati equation of stable factor matrices (Manousiouthakis 1989) we show that if the structure of the decentralized controller $K_d(s)$ is based on observers, then the form of the parameter $Q_d(s)$ can be determined in closed form

Theorem 1 The set of all concurrent decentralized controller (7) is parametrized by

$$Q_d(s) \leftrightarrow \left[\begin{array}{cc|c} A_F & BF & \bar{L} \\ 0 & A_L & \hat{L} - L \\ \hline \bar{F} - F & \bar{F} & \end{array}\right] \in \mathscr{H} \tag{13}$$

where

$$\bar{C} = \begin{bmatrix} C_1 & 0 \\ & C_2 \end{bmatrix}$$

$$\bar{L} = \begin{bmatrix} \bar{L}_1 & 0 \\ 0 & L_2 \end{bmatrix} \quad \hat{F} = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}$$

$$A_F = A - BF \tag{14}$$

$$A_L = \begin{bmatrix} A - B_2F_2 & B_2F \\ B_1F_1 & A - B_1F_1 \end{bmatrix} - \bar{L}\bar{C}$$

The proof of Theorem 1 is given in Appendix A Note that the closed loop stability of system (1) controlled by (7) is guaranteed if $\hat{A} = A$ and $Q_d(s) \in \mathscr{H}$ that is $\mathrm{Re}\,(\lambda(A_F)) < 0$ and $\mathrm{Re}\,(\lambda(A_L)) < 0$ are sufficient to guarantee the closed loop stability This will be shown for the H_2 control problem in the next section In the H_∞ control problem $\hat{A} \neq A$ such that $Q_d(s) \in \mathscr{H}$ no longer guarantees stability unless $\|Q_d(s)\|_\infty$ is less than a certain bound

A similar parameter $Q_d(s)$ can be obtained for other observer controller structures such as the sequential design in Gundeş and Desoer (1990)

3 A two stage H_2 *decentralized controller design*

In this section we will investigate the design problem

$$\min_{K_d(s) \in \mathscr{K}_d} \|T_{zw}\|_2 \tag{15}$$

for the system (1) using the concurrent decentralized observer controller $K_d(s)$ (7)

The first stage of the decentralized controller design requires an optimal H_2 centralized design Using the standard notation (Doyle *et al* 1989) let the Hamiltonian matrix have the the form

$$H = \begin{bmatrix} A & -BB^T \\ -C^TC & A^T \end{bmatrix} \tag{16}$$

where (A, B) is stabilizable and (A, C) is detectable Then H is in the domain of Riccati Operator denoted by dom (Ric) and $W = \mathrm{Ric}\,(H) \geq 0$ solves the Riccati equation

$$A^TW + WA - WBB^TW + C^TC = 0 \tag{17}$$

It follows from Assumption 1 that

$$H_2 = \begin{bmatrix} A & -BB^T \\ -C_z^TC_z & -A^T \end{bmatrix}, \quad J_2 = \begin{bmatrix} A^T & -C^TC \\ -B_wB_w^T & -A \end{bmatrix} \tag{18}$$

belong to dom (Ric) Define $X_2 = \mathrm{Ric}\,(H_2)$ and $Y_2 = \mathrm{Ric}\,(J_2)$

For system (1), the unique optimal controller that minimizes $\|T_{zw}\|_2$ is (Doyle *et al* 1989)

$$K^*(s) \leftrightarrow \left[\begin{array}{c|c} A_2 & L \\ \hline -F & 0 \end{array}\right] \tag{19}$$

where

$$F = B^TX_2, \quad L = Y_2C^T \quad A_2 = A - BF - LC \tag{20}$$

Furthermore, the optimal centralized cost is

$$\epsilon = \min \|T_{zw}\|_2 = \|T_{zw}(K^*(s))\|_2. \tag{21}$$

The second stage of the design, consisting of optimizing over the set of decentralizing parameters $\mathcal{Q}_d$, is stated in the following theorem.

Theorem 2 (optimality of concurrent decentralized observer controllers).

(1) The set of all concurrent decentralized stabilizing observer controllers (7) is given by $Q_d(s)$ (13) where $\bar{A} = A$, F and L are optimal centralized feedback and observer gains (20), $\bar{F}_1$, $\bar{F}_2$ are the feedback gains, and $\bar{L}_1$, $\bar{L}_2$ are the decentralized observer gains from (7).

(2) If $\bar{F}$ and $\bar{L}$ are chosen such that the resulting parameter $Q_d^*(s)$ achieves

$$\delta_2^* = \min_{Q_d(s) \in \mathcal{Q}_d} \|Q_d(s)\|_2 = \|Q_d^*(s)\|_2, \tag{22}$$

then the controller

$$K_d(s) = \left[\begin{array}{cc|cc} A - B\hat{F} - \bar{L}_1 C_1 & 0 & \bar{L}_1 & 0 \\ 0 & A - B\hat{F} - L_2 C_2 & 0 & \bar{L}_2 \\ \hline -\bar{F}_1 & 0 & 0 & 0 \\ 0 & -\bar{F}_2 & 0 & 0 \end{array}\right] \tag{23}$$

is the optimal decentralized observer controller which minimizes (15) and achieves the optimal cost $\alpha_2^* = \sqrt{\epsilon^2 + (\delta_2^*)^2}$.

Proof. Part 1 of Theorem 2 follows from Theorem 1. The stabilizing property of the controller is shown in Appendix B. Part 2 follows from Theorem 2 of Doyle *et al.* (1989) which when applied to the decentralizing parameter $Q_d(s)$, results in

$$\|T_{zw}^d\|_2 < \alpha_2, \tag{24}$$

if and only if

$$\|Q_d(s)\|_2^2 = \delta_2^2 < \alpha_2^2 - \epsilon^2. \tag{25}$$

□

From Theorem 2, we can interpret δ_2 as the cost of decentralization, that is, the additional cost that will be incurred if a decentralized observer controller is used instead of the optimal centralized controller.

To find the gains $\bar{F}$ and L, we propose a state space optimization procedure. Define

$$\tilde{A} = \begin{bmatrix} \bar{A}_F & B\bar{F} \\ 0 & \bar{A} \end{bmatrix} \quad \tilde{F} = [F - \hat{F} \quad F], \quad \tilde{L} = \begin{bmatrix} L \\ \bar{L} - L \end{bmatrix}, \tag{26}$$

such that

$$Q_d(s) = \tag{27}$$

Then

$$\min_{Q_d(s) \in \mathcal{Q}_d} \|Q_d(s)\|_2 = \min_{\bar{F}\,\bar{L}} \tag{28}$$

From the properties of H_2 norms (Doyle *et al.*, 1989) we obtain

$$\|Q_d(s)\|_2^2 = \mathrm{tr}\,(\tilde{F} Y \tilde{F}^T), \tag{29}$$

where

$$\tilde{A} Y + Y \tilde{A}^T + \tilde{L}\tilde{L}^T = 0. \tag{30}$$

We cast the problem of minimizing (29) subject to (30) as the minimization of the Lagrangian

$$\mathcal{L} = \mathrm{tr}\,(\tilde{F} Y \tilde{F}^T) + \mathrm{tr}\,((\tilde{A} Y + Y \tilde{A}^T + \tilde{L}\tilde{L}^T)P), \tag{31}$$

with respect to $\bar{F}$ and $\bar{L}$, where P is the (symmetric) matrix of Lagrange mulipliers. Setting the partial derivatives of (31) with respect to Y, P, $\bar{F}_1$, $\bar{F}_2$, $\bar{L}_1$, and $\bar{L}_2$ to zero, we obtain

$$\frac{\partial \mathcal{L}}{\partial P} = \tilde{A} Y + Y \tilde{A}^T + \tilde{L}\tilde{L}^T = 0, \tag{32}$$

$$\frac{\partial \mathcal{L}}{\partial Y} = \tilde{A}^T P + P \tilde{A} + \tilde{F}^T \tilde{F} = 0, \tag{33}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \bar{F}_1} = \bar{F}_1(& Y_4 - Y_2 - Y_2^T) + F_1(Y_2 - Y_1) \\ & + B_1^T(P_1(Y_2 - Y_1) + P_2(Y_4 - Y_2^T) + P_3(Y_5^T - Y_3^T) \\ & + P_3^T(Y_2 - Y_3) + P_5^T(Y_4^T - Y_5) + P_6(Y_5^T - Y_6)) \\ = 0, & \end{aligned} \tag{34}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \bar{F}_2} = \bar{F}_2(Y_1 + & Y_6 - Y_3 - Y_3^T) + F_2(Y_3 - Y_1) \\ & + B_2^T P_1(Y_3 - Y_1) + P_2(Y_5 - Y_2^T) + P_3(Y_6^T - Y_3^T) \\ & - P_2^T(Y_2 - Y_3) - P_4^T(Y_4^T - Y_5) - P_5(Y_5^T - Y_6)) \\ = 0, & \end{aligned} \tag{35}$$

$$\begin{aligned} = (P_2^T + P_4 + P_5)L_1 + (P_2^T Y_2 + P_4 Y_4 + P_5 Y_5^T) \\ \times C_1^T - P_4 \bar{L}_1 = 0, \end{aligned} \tag{36}$$

$$\begin{aligned} \partial \mathcal{L} = (P_3^T + P_5^T + P_6)L_2 + (P_3^T Y_2 + P_5^T Y_5 + P_6 Y_6) \\ \times C_2^T - P_6 \bar{L}_2 = 0, \end{aligned} \tag{37}$$

where

$$Y = \begin{bmatrix} Y_1 & Y_2 & Y_3 \\ Y_2^T & Y_4 & Y_5 \\ Y_3^T & Y_5^T & Y_6 \end{bmatrix}, \quad \begin{bmatrix} P_1 & P_2 & P_3 \\ P_2^T & P_4 & P_5 \\ P_3^T & P_5^T & P_6 \end{bmatrix} \tag{38}$$

The Lagrangian $\mathcal{L}$ is locally convex, and (32)–(37) are the necessary conditions for a local optimum. Equations (32)–(37) are coupled algebraic equations, and have no closed form solutions. These equations can be solved numerically using a steepest descent method such as the descent Anderson–Moore scheme (Mäkilä and Toivonen, 1987) or a Newton's method such as the feedback descent scheme (Beseler *et al.*, 1992). These techniques must be initiated by a stabilizing controller. In addition, the stabilizing property must be maintained during the iterative process, which can be accomplished by allowing stepsizes that preserve stability

We note that although the decentralized observer controllers implement full state feedback in A_F, the optimal feedback gains $\bar{F}_1$ and $\bar{F}_2$ are not necessarily the same as those of the optimal centralized gains F_1 and F_2. This can be attributed to the increased observation cost due to decentralization which would affect the design of $\bar{F}_1$ and $\bar{F}_2$. However, if it is determined that setting $\bar{F}_1 = F_1$ and $\bar{F}_2 = F_2$ does not significantly increase the cost, the optimization (28) reduces to

$$\min_{\bar{L}} \quad A_L \quad \bar{L} - L \tag{39}$$

which is a lower dimensional optimization problem. A set of necessary conditions of dimensions lower than those of (32)–(37) can be derived

Example. To illustrate the H_2 decentralized design, we present an example of optimal decentralized control system in Fig. 2 to achieve tracking of a reference signal. Let the plant model be

$$P(s) = \begin{matrix} & 1 & \\ & -2 & \\ & 0 & 0 \\ & 0 & 0 \end{matrix} \tag{40}$$

To track step input at r_1, we represent the inputs as a weighted disturbance, and the tracking error as an error

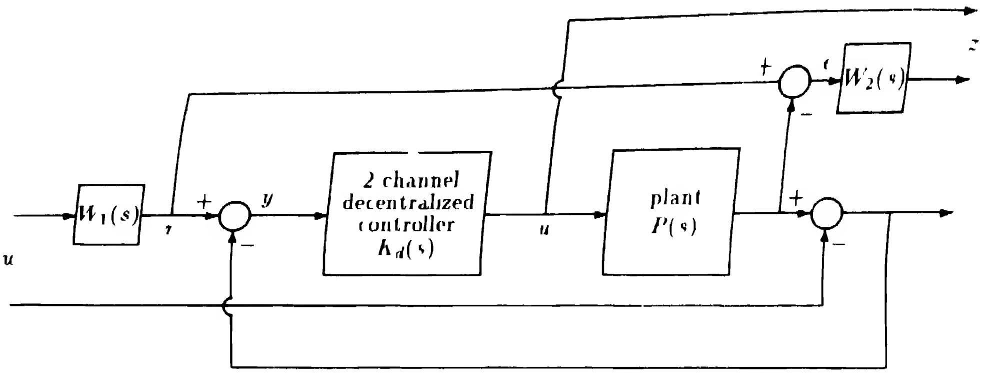


FIG. 2. Control system for H_2 optimization.

output. The weighting functions W_1 and W_2 are chosen as

$$W_1 \leftrightarrow \left[\begin{array}{c|cc} -0.001 & 10 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \qquad W_2 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix} \tag{41}$$

to approximate the step inputs at r_1, and to penalize the tracking error more heavily than other error outputs.

The matrices for the generalized plant (1) are

$$A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -2 & 0 \\ 0 & 0 & -0.001 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$B_w = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 \end{bmatrix}, \quad C_z = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{42}$$

$$D_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_{21} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For this system the centralized H_2 optimal cost is 458.19.

The H_2 optimal decentralized observer controller for system (1) with the data (42) is designed using the feedback descent algorithm (Beseler *et al.* 1992). The controller is found to be

$$K_d(s) = \begin{bmatrix} \dfrac{999.2(s+98.79)(s+0.9288)}{(s+109.97)(s+98.712)(s+0.0027)} & 0 \\ 0 & \dfrac{-291.46(s+97.28)(s+3.48)}{(s+103.86)(s+97.7)(s+0.002)} \end{bmatrix} \tag{43}$$

and the optimal decentralized H_2 cost and the cost of decentralization are, respectively

$$\|T^d_{zw}\|_2 = 489.39 \quad \delta = 171.94 \tag{44}$$

Figure 3 illustrates the tracking of a step in channel 1 by the resulting control system. The performance of the centralized optimal controller is also shown for comparison. We note that the tracking response with the decentralized control closely approximates that of the optimal centralized controller and there is no serious loss of performance.

It is of interest to note that the decentralized F_1, F are very close to the centralized F, F. If we impose F_1, F_1 and $F_2 = F_2$ the H_2 cost and the cost of decentralization achieved by the resulting controller match those of (44) to five significant digits.

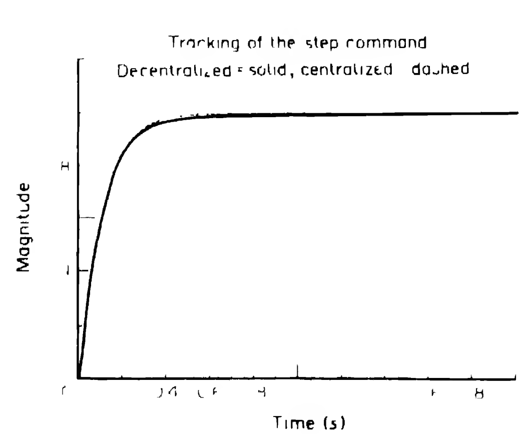


FIG. 3. Tracking response of H_2 optimal decentralized controller.

4. A two-stage H_∞ decentralized controller design

In this section we will investigate the decentralized H_∞ suboptimal control problem of finding a concurrent decentralized observer controller such that

$$\|T^d_{zw}\|_\infty < \gamma \tag{45}$$

where γ is a pre-specified constant. As in the H_2 design the H_∞ decentralized design also involves two stages—a centralized suboptimal design in the first stage and the optimization of the decentralizing parameter Q_d in the second stage.

The suboptimal H_∞ design in Doyle *et al.* (1989) will be used in the centralized design. From Assumption 1 there exists a $\gamma > 0$ such that

$$H_\infty = \begin{bmatrix} A & \gamma^{-2}B_wB_w' - BB' \\ -C_z'C_z & -A' \end{bmatrix}$$
$$J_\infty = \begin{bmatrix} A' & \gamma^{-2}C_z'C_z - C'C \\ -B_wB_w' & -A \end{bmatrix} \tag{46}$$

belong to dom(Ric). If $X_\infty = \mathrm{Ric}(H_\infty) \geq 0$, $Y_\infty = \mathrm{Ric}(J_\infty) \geq 0$ and $\rho(X_\infty Y_\infty) < \gamma^2$, where $\rho(\cdot)$ denotes the spectral radius, then a controller that achieves $\|T_{zw}\|_\infty < \gamma$ is

$$K_{sub}(s) \leftrightarrow \left[\begin{array}{c|c} A_\infty & L_\infty \\ \hline F_\infty & 0 \end{array}\right] \tag{47}$$

where

$$A_\infty = A + \gamma^{-2}B_wB_w'X_\infty - BF_\infty - Z_\infty L_\infty C,$$
$$F_\infty = B'X_\infty, \quad L_\infty = Y_\infty C', \quad Z_\infty = (I - \gamma^{-2}Y_\infty X_\infty)^{-1} \tag{48}$$

Given the centralized controller (47) achieving $\|T_{zw}\|_\infty < \gamma$, the second stage of the design process is to solve for the decentralizing parameter having the property

$$\|Q_d(s)\|_\infty < \gamma. \tag{49}$$

Theorem 3 (suboptimality of concurrent decentralizer observer controllers).

(1) The set of all concurrent decentralized observer controllers (7) is given by $Q_d(s)$ (13) where

$$\hat{A} = A_{\rm tmp} = A + \gamma^{-2}B_wB_w^TX_\infty, \tag{50}$$

$F = F_\infty$ and $L = L_\infty$ are the suboptimal centralized feedback and observer gains (48), X_∞ is given by (46), $\hat{F}_1, \hat{F}_2$ are the feedback gains, and $\hat{L}_1, \hat{L}_2$ are the decentralized observer gains from (7)

(2) If $\hat{F}$ and $\hat{L}$ are chosen such that the resulting parameter $Q_d(s)$ achieves

$$\|Q_d(s)\|_\infty < \gamma, \tag{51}$$

then the controller

$$K_d(s) \cdot \left[\begin{array}{cc|cc} A_{\rm tmp} - B\hat{F} - \hat{L}_1C_1 & 0 & \hat{L}_1 & 0 \\ 0 & A_{\rm tmp} - B\hat{F} - \hat{L}_2C_2 & 0 & \hat{L}_2 \\ \hline -\hat{F}_1 & 0 & 0 & 0 \\ 0 & -\hat{F}_2 & 0 & 0 \end{array}\right] \tag{52}$$

is a suboptimal stabilizing decentralized observer controller satisfying $\|T^d_{zw}\|_\infty < \gamma$.

Proof. Part 1 of Theorem 3 follows from Theorem 1 Part 2 follows from Theorem 4 of Doyle *et al*. (1989), which also contains the result that the control $K_d(s)$ (52) is a stabilizing control if (51) is satisfied. □

Theorem 3 determines whether the solution to a decentralized control design exists for a given γ Theorem 3 suggests an iterative two-stage procedure to compute the optimal decentralized H_∞ problem for the concurrent controller structure, that is, the minimum γ admitting a decentralized solution. At each stage of the iterative process, the existence of a decentralized controller for a trial γ is determined. Then a bisection technique can be used to find the smallest γ that admits a decentralized controller. Then a cost of decentralization can be found as the difference between the minimum γ for the centralized control and the decentralized control.

We now propose a state space approach to find the gains $\hat{F}$ and $\hat{L}$ to solve for (51). Defining

$$\begin{bmatrix} \hat{A}_F & BF \\ 0 & \hat{A}_l \end{bmatrix} \quad \bar{F} = [F - \hat{F} \quad \hat{F}], \quad \bar{L} = \begin{bmatrix} L \\ \hat{L} - L \end{bmatrix}, \tag{53}$$

such that

$$Q_d(s) \leftrightarrow \tag{54}$$

the synthesis of $Q_d(s)$ (7) subject to (51) requires the existence of $W \geq 0$ to satisfy the Riccati equation

$$\bar{A}^TW + W\bar{A} + \gamma^{-2}W\bar{L}\bar{L}^TW + \bar{F}^T\bar{F} = 0. \tag{55}$$

Instead of directly attempting to solve (55), we propose a simplification by assuming $\hat{F}_1 = F_1$ and $\hat{F}_2 = F_2$ where

$$= \begin{bmatrix} F_1 \\ F_2 \end{bmatrix},$$

is the centralized suboptimal gain (48). As a result, $Q_d(s)$ simplifies to

$$Q_d(s) \cdot \quad \hat{A}_l \quad \hat{L} - \bar{L}$$

Thus we are required to design $\hat{L}$ such that the Riccati equation

$$\bar{A}_l^TM + M\bar{A}_l + \gamma^{-2}M(\hat{L} - \bar{L})(\hat{L} - \bar{L})^TM + \bar{F}^T\bar{F} = 0, \tag{58}$$

admits a solution $M \geq 0$. Without loss of generality we consider $M > 0$ solutions since the uncontrollable and unobservable parts of $Q_d(s)$ for all selection of $\hat{L}$ can be eliminated.

Let $Z = \gamma^2M^{-1}$. Then premultiplying (58) by Z and postmultiplying by M^{-1} and using the expression of $\bar{A}_l$ from (14) we obtain

$$0 = Z(\bar{A} - \bar{L}\bar{C})^T + (\bar{A} - \bar{L}\bar{C})Z + \gamma^{-2}Z\bar{F}^T\bar{F}Z - Z\bar{C}^T\bar{C}Z + (\hat{L} - \bar{L} + Z\bar{C}^T)(\hat{L} - \bar{L} + Z\bar{C}^T)^T. \tag{59}$$

In the absence of the decentralized constraint $\hat{L} = \mathrm{diag}(\hat{L}_1, \hat{L}_2)$, we could choose $\hat{L} = \bar{L} + Z\bar{C}^T$ to eliminate the last term of (59), and convert the resulting equation into a Lyapunov equation which has a direct solution. Under the decentralized constraint, only the diagonal blocks can be equated. Following an idea in Veillette (1990) we let

$$\hat{L}_1 = L_1 + Z_{11}C_1^T, \quad \hat{L}_2 = L_2 + Z_{22}C_2^T, \tag{60}$$

where

$$Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}. \tag{61}$$

Since Z is symmetric, (59) can further be written as

$$0 = Z(\bar{A} - \bar{L}\bar{C})^T + (\bar{A} - \bar{L}\bar{C})Z + \gamma^{-2}Z\bar{F}^T\bar{F}Z - Z\bar{C}^T\bar{C}Z \cdot \begin{bmatrix} 0 & L_2 + Z_{12}C_2^T \\ L_1 + Z_{21}C_1^T & 0 \end{bmatrix}\begin{bmatrix} 0 & L_2 + Z_{12}C_2^T \\ L_1 + Z_{21}C_1^T & 0 \end{bmatrix}^T. \tag{62}$$

From this derivation, we establish the following result.

Theorem 4. Let the gains F_1, F_2, L_1 and L_2 be obtained from (48) for a given γ and set $\hat{F}_1 = F_1$ and $\hat{F}_2 = F_2$. If there exists a $Z > 0$ satisfying (62), then the observer gains (60) guarantee that $\|Q_d(s)\|_\infty < \gamma$, and the decentralized observer-based controller (7) achieves $\|T^d_{zw}\|_\infty < \gamma$.

Theorem 4 provides a means to compute a decentralized controller for a specific γ. Although (62) cannot be solved directly, an algorithm can be designed to iteratively solve for Z. To initiate the algorithm, assume $Z = 0$ and compute the last term of (62). Then solve (62) as a Riccati equation for Z. Update the last term of (62) and repeat the solution process until the difference between successive solutions of Z is small. The solution process can proceed provided that $Z > 0$. In contrast to the H_2 algorithm for solving (32)–(37), the iterative solution of (62) does not require an initial stabilizing $\hat{H}$. An equation similar to (62) has been derived in Veillette (1990) for decentralized H_∞ control design. In Veillette (1990), the design equation was obtained based on the closed-loop transfer function, and hence, contains terms different from those in (62).

Example We present an H_∞ suboptimal controller design for attenuation of the measurement noise for the plant

$$P(s) \cdot \left[\begin{array}{ccc|cc} -1 & 1 & 1 & 0 & 0 \\ 0 & -5 & 0 & 1 & 0 \\ 0 & 0 & -10 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{array}\right] \tag{63}$$

To weigh the plant output errors more heavily than the control signal, we weigh the error signals by 50.

The matrices for the generalized plant (1) are

$$A = \begin{bmatrix} -1 & 1 & 1 \\ 0 & -5 & 0 \\ 0 & 0 & -10 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad C_z = \begin{bmatrix} 50 & 50 & 0 \\ 50 & 0 & 50 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$D_{12} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_{21} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{64}$$

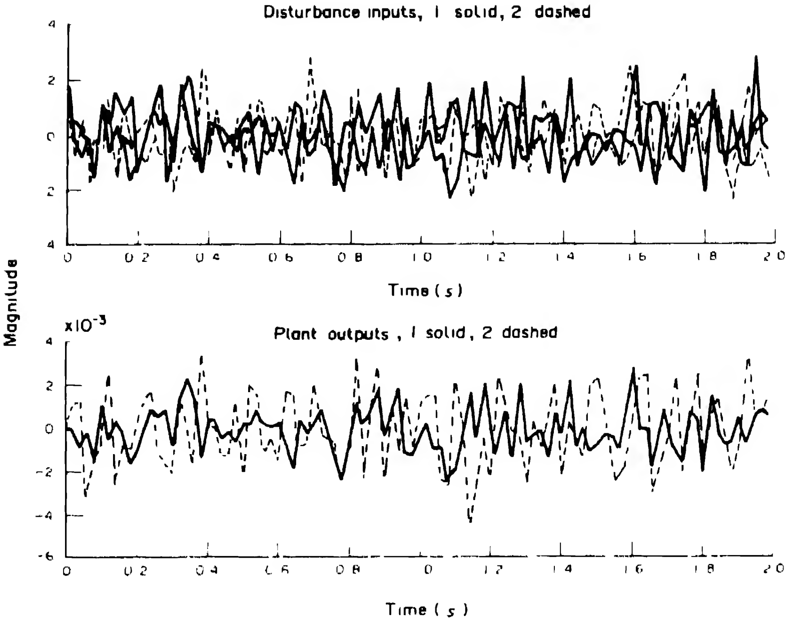


FIG. 4. Performance of the H_∞ suboptimal controller.

We impose a bound of $\gamma = 1$, implying that for disturbances with a unit magnitude, the output will be guaranteed to be less than $\gamma/50 = 0.02$. The controller designed using the proposed iterative algorithm to solve (62) is

$$K_d(s) = \begin{bmatrix} \dfrac{-0.865}{(s + 890.12)(s + 295.68)(s + 4.443)} & 0 \\ 0 & \dfrac{-0.712}{(s + 890.12)(s + 295.68)(s + 4.443)} \end{bmatrix} \qquad (65)$$

Figure 4 verifies that the design objectives have been satisfied, resulting in an attenuation of 10^{-3} (60 dB).

5. *Conclusions*

A two-stage design approach to optimal H_2 and H_∞ decentralized control problems has been proposed and applied to the design of concurrent decentralized observer controllers. The first stage design is a standard optimal centralized controller design. The second stage involves the optimization of the decentralizing parameter and is, in general, non-trivial for high-dimensional systems. The parametrization approach offers some simplification for the computation of suboptimal decentralized controllers.

Further work in this area includes investigating the optimal H_2 and H_∞ decentralized control problems using other decentralized controller parametrizations. One of the possible parametrizations is contained in Date (1991) where a decentralized controller is used as the central controller in the parametrization. It will be of interest to see whether simplified design algorithms can be developed.

Acknowledgements—The authors wish to thank Dr Dean Minto, Dr R. Ravi and Dr Bob Veillette for helpful discussions.

References

Bernstein, D. S. (1987). Sequential design of decentralized dynamic compensators using optimal projection equations. *Int. J. Control*, **46**, 1569–1577.

Beseler, J., J. H. Chow and K. D. Minto (1992). A feedback descent method for solving constrained LQG control problems. *Proc.* 1992 *ACC*, pp. 1044–1048.

Date, R. A. (1991). Decentralized control: stable factorization approach to parametrization and design. Ph.D thesis, Electrical, Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, New York.

Date, R. A. and J. H. Chow (1989). A reliable coordinated decentralized control design. *Proc. of* 1989 *CDC*, pp. 1295–1300.

Date, R. A. and J. H. Chow (1991). Decentralized stable factors and parametrization of decentralized controllers. *Proc.* 1991 *ACC*, pp. 904–909.

Davison, E. J. and W. Gesing (1979). Sequential stability and optimization of large scale decentralized control systems. *Automatica*, **15**, 307–324.

Davison, E. J. and I. J. Ferguson (1981). The design of controllers for robust servomechanism problem using parameter optimization method. *IEEE Trans. Aut. Control*, **AC-26**, 93–110.

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State space solution to standard H_2 and H_∞ control problem. *IEEE Trans. Aut. Control*, **34**, 831–846.

Francis, B. (1987). A course in H_∞ control theory. *Lecture Notes in Control and Information Sciences*, **88**. Springer-Verlag, Berlin.

Geromel, J. and J. Bernussou (1982). Optimal decentralized control of dynamic systems. *Automatica*, **18**, 545–557.

Gündes, A. N. and C. A. Desoer (1990). Algebraic theory of linear feedback systems with full and decentralized compensators. *Lecture Notes in Control and Information Sciences*, **142**. Springer-Verlag, Berlin.

Iftar, A. and Ü. Özgüner (1989). An optimal control approach to the decentralized robust servomechanism problem. *IEEE Trans. Aut. Control*, **AC-34**, 1268–1271.

Mäkilä, P. M. and H. T. Toivonen (1987). Computational methods for parametric LQ problems—a survey. *IEEE Trans. Aut. Control*, **AC-32**, 658–671.

Manousiouthakis, V. (1989). On parametrization of all decentralized stabilizing controllers. *Proc.* 1989 *ACC*, pp. 2108–2111.

Özgüler, A. B. (1990). Decentralized control: A stable proper fractional approach. *IEEE Trans. Aut. Control*, **AC-35**, 1109–1117.

Sandell, N. R., P. Varaiya, M. Athans and M. G. Safonov (1978). Survey of decentralized control methods for large scale systems. *IEEE Trans. Aut. Control*, **AC-23**, 108–128.

Siljak, D. D. (1991). *Decentralized Control of Complex Systems*. Academic Press, New York.

Wang, S. H. and E. J. Davison (1973). On stabilization of

decentralized control systems. *IEEE Trans. Aut. Control*, **AC-18**, 473–478.

Uskokovic, Z. and J. Medanic (1985) Sequential design of decentralized low order dynamic regulators. *Proc.* 1985 *CDC*, pp. 837–842.

Veillette, R. J. (1990). Reliable control of decentralized systems: an ARE-based H_∞ approach. Ph.D. thesis, Department of Electrical Engineering, Univ. of Illinois, Urbana-Champaign, IL.

Appendix A

Proof of Theorem 1 Putting $Q_d(s)$ (13) into $M(s)$ (9), we obtain the controller realization

$$K(s) \leftrightarrow \left[\begin{array}{c|c} A_m & B_m \\ \hline C_m & 0 \end{array}\right], \qquad \text{(A1)}$$

where

$$A_m = \begin{bmatrix} \hat{A} - BF - LC & BF - B\hat{F} & B_1\bar{F}_1 & B_2\bar{F}_2 \\ -LC & \hat{A} - B\hat{F} & B_1\bar{F}_1 & B_2F_2 \\ -LC + \bar{L}_1C_1 & 0 & \hat{A} - B_1\bar{F}_2 - L_1C_1 & B_2\bar{F}_2 \\ -LC + \bar{L}_2C_2 & 0 & B_1\bar{F}_1 & \hat{A} - B_1\bar{F}_1 - \bar{L}_2C_2 \end{bmatrix}, \qquad \text{(A2)}$$

$$B_m = \begin{bmatrix} L_1 & L_2 \\ L_1 & L_2 \\ L_1 - L_1 & L_2 \\ L_1 & L_2 - \bar{L}_2 \end{bmatrix}, \qquad \text{(A3)}$$

$$C_m = \begin{bmatrix} -F_1 & F_1 - \bar{F}_1 & F_1 & 0 \\ -F_2 & F_2 - \bar{F}_2 & 0 & F_2 \end{bmatrix} \qquad \text{(A4)}$$

Performing a similarity transformation with

$$T = \begin{bmatrix} I & -I & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & -I & I & 0 \\ 0 & -I & 0 & I \end{bmatrix}, \qquad \text{(A5)}$$

we obtain

$$TA_mT^{-1} = \begin{bmatrix} \hat{A} - BF & 0 & 0 & 0 \\ -LC & \hat{A} - B\hat{F} & 0 & 0 \\ \bar{L}_1C_1 & 0 & \hat{A} - B\bar{F} - \bar{L}_1C_1 & 0 \\ \bar{L}_2C_2 & 0 & 0 & \hat{A} - B\hat{F} - \bar{L}_2C_2 \end{bmatrix}, \qquad \text{(A6)}$$

$$TB_m = \begin{bmatrix} 0 & 0 \\ L_1 & L_2 \\ -\bar{L}_1 & 0 \\ 0 & -\bar{L}_2 \end{bmatrix}, \qquad \text{(A7)}$$

$$C_mT^{-1} = \begin{bmatrix} -F_1 & 0 & \bar{F}_1 & 0 \\ -F_2 & 0 & 0 & \bar{F}_2 \end{bmatrix}. \qquad \text{(A8)}$$

Note that the eigenvalues of $\hat{A} - BF$ are not controllable and those of $\hat{A} - B\bar{F}$ are not observable. The elimination of these eigenvalues reduces the controller (A1) to (7).

Appendix B

Proof of the stabilizing property of the decentralized observer controller in Theorem 2.

With $\hat{A} = A$, the closed-loop poles of the system (1) controlled by the decentralized observer controller (7) are the eigenvalues of the matrix

$$A_c = \begin{bmatrix} A & -B_1\bar{F}_1 & -B_2\bar{F}_2 \\ \bar{L}_1C_1 & A - B\bar{F} - \bar{L}_1C_1 & 0 \\ \bar{L}_2C_2 & 0 & A - B\bar{F} - \bar{L}_2C_2 \end{bmatrix} \qquad \text{(B1)}$$

Performing a similarity transformation with

$$T_c = \begin{bmatrix} I & 0 & 0 \\ -I & I & 0 \\ -I & 0 & I \end{bmatrix}, \qquad \text{(B2)}$$

we obtain

$$T_cA_cT_c^{-1} = \begin{bmatrix} \bar{A}_f & [-B_1\bar{F}_1 - B_2\bar{F}_2] \\ 0 & \bar{A}_l \end{bmatrix}, \qquad \text{(B3)}$$

which has stable eigenvalues if $Q_d(s) \in \mathcal{Q}_d$

Automatica, Vol 29, No 2, pp 465–467, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

On Pole Placement with Decentralized Static Feedback*

P. FESSAS†

Key Words—Pole placement; interconnected systems; local state-vector feedback; decentralized control.

Abstract—This note contains an investigation of the pole placement problem in interconnected systems (A, B_d), when local, static state-vector feedbacks are applied. It is examined only for systems consisting of two interconnected subsystems, and it is shown that it has a solution under the very general assumptions of the global and the local controllability, when additionally system matrix A has a particular structure. It is also proven that, with the present approach, at least one eigenvalue of the closed-loop system has to be real, and satisfy a certain condition.

1. *Introduction*

IN THIS NOTE the decentralized pole placement problem with static feedback of the local state vectors in an interconnected system is considered. It is examined for an interconnected (global) system consisting of two interconnected (local) scalar subsystems, under the very general assumptions of the global and the local controllability. Only the case of two interconnected systems is examined, since only then will the global system have no decentralized fixed modes (Wang and Davison, 1974; Davison and Ozguner, 1983; Anderson and Clements, 1979; Calogiannis and Fessas, 1982; Fessas, 1982). It is shown that this problem admits a solution, up to the restriction that at least one of the eigenvalues of the closed-loop system is real, when, additionally, the system matrix is of a particular structure (corresponding to the rank $[G] = 1$ case; the notation used is explained below). It is also indicated that, when the system matrix has no particular structure, the same result holds. A more general result concerning only the stabilizability with static feedbacks of the local state-vectors of a system consisting of two interconnected subsystems was earlier proven (Fessas, 1987). The main contribution of the present paper is to show—additionally—that with such a feedback $(n-1)$ eigenvalues of the nth order interconnected system can be arbitrarily placed. It is obvious that the present result includes the former one.

2. *Preliminaries*

2.1. *Form of system* (A, B_d). We consider the interconnected system (A, B_d) defined by

$$\dot{x} = Ax + B_d u, \tag{2.1}$$

where x is the n-dimensional state of (A, B_d), u is its two-dimensional input vector, A is the n by n system matrix, and B_d is its $n \times 2$ input matrix. Matrices A and B_d admit the following partitioning

$$A \quad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{matrix} \updownarrow n_1 \\ \updownarrow n_2 \end{matrix} \quad B_d = \begin{bmatrix} & \\ 0 & b_{22} \end{bmatrix}, \tag{2.2}$$

with $n = n_1 + n_2$. System (A, B_d) consists of the interconnected n_i-dimensional subsystems (A_{ii}, b_{ii}) – $i = 1, 2$ – of local state-vectors x_1 and x_2, with $x = [x_1' \quad x_2']'$, u_1 and u_2 being, respectively the scalar inputs of these subsystems, with $u = [u_1 \quad u_2]'$. Without loss of generality (Davison and Ozguner, 1983), we assume that these two inputs are scalar. We further assume that the global system (A, B_d), as well as its two subsystems (A_{ii}, b_{ii})—$i = 1, 2$—are controllable. In that case, subsystems (A_{ii}, b_{ii}) are supposed to be in their companion controllable form (Kailath, 1980):

$$\begin{matrix} 0 & 1 \\ & \\ & a_{ii}' \end{matrix} \tag{2.3}$$

where a_{ii}' denotes the last row elements of A_{ii}. When (A_{ii}, b_{ii}) are in the above form, the $n_i \times n_j$ submatrices A_{ij} $(i \neq j)$ of A assume no particular form: for them we use the notation:

$$A_{ij} = \begin{bmatrix} A_{ij}^0 \\ a_{ij}' \end{bmatrix}, \quad (i \neq j), \tag{2.4}$$

where a_{ij}' denotes the last row elements of A_{ij}, and A_{ij}^0 the others. It is obvious that when the various submatrices of (A, B_d) are in the above form, system (A, B_d) is as:

$$A = \left.\begin{array}{ccc:ccc} 0 & 1 & & & A_{12}^0 & \\ & & \ddots & & & \\ & & 1 & & & \\ & a_{11}' & & & a_{12}' & \\ \hdashline & A_{21}^0 & & 0 & 1 & \\ & & & & \ddots & \\ & & & & & 1 \\ & a_{21}' & & & a_{22}' & \end{array}\right] , \quad B_d = \quad 0$$

Finally, with the elements of rows n_1 and $n_1 + n_2 (= n)$ of A, we form matrix A_m:

$$\cdot \begin{bmatrix} a_{11}' & a_{12}' \\ a_{21}' & a_{22}' \end{bmatrix}. \tag{2.5}$$

2.2. *The intercontrollability matrix* D(s) *and its kernel* U(s). The following $(n-2) \times n$ polynomial matrix is the

* Received 31 October 1991; revised 14 June 1992; received in final form 5 September 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor T. Başar under the direction of Editor A. P. Sage.

† Department of Electrical Engineering, Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece.

intercontrollability matrix of system $(A\ B_d)$

$$D(s) \quad \begin{matrix} s & -1 & & & \\ & & & & -A^0_{1,} \\ & & s & -1 & \\ & & & s & -1 \\ -A^0_{,1} & & & & \\ & & & s & -1 \end{matrix} \qquad (2\ 6)$$

As the following lemma indicates $D(s)$ expresses the conditions for the controllability of $(A\ B_d)$

Lemma 2 1 (Calogiannis and Fessas 1982) System $(A\ B_d)$ is controllable if and only if rank $D(s) = n - 2$ Thus the matrix $D(s)$ of a controllable system $(A\ B_d)$ is a full rank matrix Its kernel $U(s)$ is an $n \times 2$ polynomial matrix of rank 2 such that $D(s)U(s) = 0$

The analytical determination of $U(s)$ is as follows P is the matrix representing the column permutations of matrix $D(s)$ which brings it to the form of the matrix pencil $D(s)$ with

$$D(s) = D(s)P \quad [sI_{n-2} \quad F \mid G] \qquad (2\ 7)$$

In (2 7) G is an $(n-2) \times 2$ (constant) matrix consisting of columns n_1 and $n_1 + n_2 - n$ of $D(s)$ F is an $(n\ \ 2) \times (n\ \ 2)$ constant matrix and I_{n-2} is the unity matrix of order $n - 2$ Since $D(s)$ is a full rank matrix the pair $(F\ G)$ is controllable In case G has the property rank $[G] = 1$ with $\varepsilon_1 g_1 + \varepsilon_2 g_2 = 0$ for some non zero numbers ε_1 and ε_2 F must be a cyclic matrix with the characteristic polynomial $X(s)$ The pair $(F\ G)$ is transformed into its companion controllable form by the similarity transformation T The polynomial vector $S_{n-2}(s) = [1\ s \quad s^{n-3}]$ is the structure operator of $(F\ G)$ The form of $U(s)$ is the content of the following lemma

Lemma 2 2 Let $D(s)$ be the intercontrollability matrix of $(A\ B_d)$ as in (2 1) with rank $[G] = 1$ for G as in (2 7) Then the kernel $U(s)$ of $D(s)$ is equal to

$$U(s) = P \begin{bmatrix} TS_{n-2}(s) & 0 \\ -X(s) & \varepsilon_1 \\ 0 & \varepsilon_2 \end{bmatrix} \qquad (2\ 8)$$

where $P\ T\ S_{n-2}(s)\ X(s)\ \varepsilon_1$ and ε_2 are as previously explained

3 *Main development*

Consider the interconnected system $(A\ B_d)$ with A and B_d as in (2 2) and (2 3) In that case the corresponding differential equation in the state space is

$$\dot{x}(t) = Ax(t) + B_d u(t) \qquad (3\ 1)$$

In the operator domain to this equation corresponds the equation

$$(sI - A)x(s) = B_d u(s) \qquad (3\ 2)$$

which in its turn reduces to the equations

$$D(s)x(s) = 0 \qquad (3\ 3a)$$

and

$$(sE - A_m)x(s) = u(s) \qquad (3\ 3b)$$

$D(s)$ is as in (2 6) E is a $2 \times n$ (constant) matrix of the form $E = \mathrm{diag}\{e_1 e_2\}$ the n_i dimensional vector e_i being equal to $e_i = [0 \quad 0\ 1]$—for $i = 1\ 2$—and A_m is the matrix defined in (2 5) From (3 3a) it follows that $x(s)$ must satisfy the relation

$$x(s) = U(s)\xi(s) \qquad (3\ 4)$$

where $U(s)$ is the kernel of $D(s)$ and $\xi(s)$ is any two-dimensional vector It follows that $\xi(s)$ must satisfy the equation

$$M(s)\xi(s) = u(s) \qquad (3\ 5)$$

The matrix $M(s)$ appearing in (3 5) is termed characteristic matrix of the interconnected system (A, B_d) (Fessas, 1988) and is defined by the relation

$$M(s) = (sE - A_m)U(s) \qquad (3\ 6)$$

The three systems defined, respectively (i) in the state-space by the pair of matrices $(A\ B_d)$, (ii) in the operator domain by $\{sI - A\ B_d\}$ (iii) by the polynomial matrix description (PMD)

$$M(D)\xi(t) = u(t) \qquad (3\ 7a)$$

$$x(t) = U(D)\xi(t) \qquad (3\ 7b)$$

are equivalent (Wolovich 1974, Kailath 1980, Chen, 1984) It is noted that in (3 7) $\xi(t)$ is the pseudo-state vector of the system and is related to the state vector $x(t)$ of system $(A\ B_d)$ by the relation

$$x(t) = U(D)\xi(t) \qquad (3\ 8)$$

(in (3 7) (3 8) the symbol D denotes the differential operator d/dt)

We now consider the local feedbacks $u_1 = K_{11}x_1$ and $u_2 = K_{22}x_2$ or $u = K_d x$ with $K_d = \mathrm{diag}\{K_{11}K_{22}\}$ This decentralized control applied to system $(A\ B_d)$ corresponds to the control $u(s) = K_d U(s)\xi(s)$ applied to the system defined by the PMD (3 7) The resulting closed loop system matrix is then

$$M_d(s) = (sE - A_m - K_d)U(s) \qquad (3\ 9)$$

We assume that the feedback matrix K_d has the specific form

$$K_d = \begin{bmatrix} \alpha_1 & \alpha_{n1} & 0 & \\ & 0 & \beta_1 & \beta_{n} \end{bmatrix} \qquad (3\ 10)$$

where $\alpha_1 \quad \alpha_{n1}\ \beta_1 \quad \beta_{n2}$ are some real numbers (the feedback coefficients) Without loss of generality we assume that $A_m = 0$ (see also the remark after the end of the proof) Then $M_d(s)$ is given by

$$\begin{matrix} & & 0 & & s \\ & & \beta_1 & -\beta_n & s - \beta_n \\ t_1(s) & & & & \\ & & \alpha(s) & (s\ \ \alpha_{n1})X(s) & \varepsilon_1(s - \alpha_{n1}) \\ & & \beta(s) & & \varepsilon_2(s - \beta_{n2}) \\ t_n(s) & & & & \\ -X(s) & & & & \\ 0 & & & & \end{matrix}$$

$$(3\ 11)$$

where $t(s) = TS_{n-2}(s) = [t_1(s) \quad (s)]$ the scalar polynomials $\{t_1(s) \quad (s)\}$ being prime not monic and

$$\alpha(s) = [\alpha_1 \quad \alpha_{n1-1} \quad 0 \quad 0]t(s) \qquad (3\ 12a)$$

$$\beta(s) = [0 \quad 0\ \beta_1 \quad]t(s) \qquad (3\ 12b)$$

Obviously $\alpha(s)$ and $\beta(s)$ are of (maximum) degree $n - 3$ The determinant of the matrix in (3 11) is the closed loop characteristic polynomial It can be (almost) arbitrarily assigned by suitable choice of the feedback coefficients $\alpha_1 \quad \alpha_{n1}$ and $\beta_1 \quad \beta_n$ The result is formulated as the following proposition

Proposition 3 1 Consider the interconnected system $(A\ B_d)$ consisting of the subsystems as in (2 2)–(2 3) and suppose that rank $[G] = 1$ with $(F\ G)$ as in (2 7) Then its $n \quad 1$ eigenvalues $\lambda_1\lambda_2 \quad \lambda_{n-1}$ can be arbitrarily assigned by the decentralized control $u = K_d x$ while the nth one—λ_n must satisfy the relation $\lambda_n = \sum_{i=1}^{n-1} \lambda_i - \sum_{j=1}^{n-2} \mu_j$ where $\mu_1\mu_2 \quad \mu_{n-2}$ are the eigenvalues of F

Proof In the matrix in (3 11) we substract the second row multiplied by the coefficient $\varepsilon_1/\varepsilon_2$ from the first and we assume that $\alpha_{n1} = \beta_{n2}(=K)$ In that case the matrix as in

(3.11) is equal to:

$$\begin{bmatrix} -\alpha(s) + \varepsilon_1/\varepsilon_2\beta(s) - (s-K)X(s) & 0 \\ -\beta(s) & \varepsilon_2(s-K) \end{bmatrix}.$$

Without loss of generality, ε_2 may be taken to equal ($\varepsilon_2 = -1$). The determinant of $M_d(s)$ is a monic polynomial of degree n, of the form: $\det M_d(s) = h(s)(s-k)$. The polynomial $h(s)$ is of degree $n-1$, and is equal to: $h(s) = \alpha(s) + \varepsilon_1\beta(s) + (s-K)X(s)$. Let

$$h(s) = (s-\lambda_1)\cdots(s-\lambda_{n-1}) = s^{n-1} + h_0 s^{n-2} + \cdots, \quad (3.13)$$

be this polynomial, corresponding to the eigenvalues $\lambda_1 \cdots \lambda_{n-1}$ and let $X(s) = s^{n-2} + x_0 s^{n-3} + \cdots = (s-\mu_1)\cdots(s-\mu_{n-2})$, (3.14) be the characteristic polynomial of system (F, G), of eigenvalues $\mu_1, \ldots, \mu_{n-2}$. The rest of the proof is as follows:

- choose $K = \lambda_n$, the nth closed-loop eigenvalue, as $K = x_0 - h_0$, where x_0 is given by (3.14), and h_0 is given by (3.13);
- compute the polynomial $[h(s) - (s-K)X(s)]$. It is noted that this polynomial is not monic, and of degree $n-3$, by the previous choice of K. Finally, determine the feedback parameters $\alpha_1 \cdots \alpha_{n1-1}, \beta_1 \cdots \beta_{n2-1}$ by equating the polynomial $\alpha(s) + \varepsilon_1\beta(s)$ to the polynomial $[h(s) - (s-K)X(s)]$.

Thus, the closed loop characteristic polynomial of (A, B_d) is $\det M_d(s) = h(s)(s-k)$, and has $\lambda_1 \cdots \lambda_{n-1}, K$ as the closed loop eigenvalues. It is also noted that since $x_0 = -\sum_{i=1}^{n-2}\mu_i$, and $h_0 = -\sum_{i=1}^{n-1}\lambda_i$, the nth eigenvalue $\lambda_n = K = \alpha_{n1} = \beta_{n2}$ is related to other eigenvalues by:

$$\lambda_n = -\sum_{j=1}^{n-2}\mu_j + \sum_{i=1}^{n-1}\lambda_i \qquad \text{Q.E.D.}$$

Remark. When $A_m \neq 0$, the proof follows exactly the same lines, with the difference that in the polynomials $\alpha(s)$ and $\beta(s)$ (as in (3.12)) appear now the various elements of A_m, except the elements (n_1, n_1), (n, n_1), (n_1, n) and (n, n) of a_m, which appear at the last two columns of matrix $(sE - A_m - K_o)P$.

Example. Let (A, B_d) be the four-dimensional system

$$A = \left[\begin{array}{cc:cc} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ \hdashline 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \quad B_d = \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \\ \hdashline 0 & 0 \\ 0 & 1 \end{array}\right].$$

The intercontrollability matrix $D(s)$ is then

$$D(s) = \begin{bmatrix} s & -1 & -1 & -1 \\ 0 & -1 & s & -1 \end{bmatrix} \quad \text{and}$$

$$\bar{D}(s) = D(s)P = \left[\begin{array}{cc:cc} s & -1 & -1 & -1 \\ 0 & s & -1 & -1 \end{array}\right],$$

with $P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. It follows that $F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and

$$G = [g_1 \quad g_2] = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}.$$

The transformation matrix T is $T = \begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix}$, the kernel $U(s)$ is

$$U(s) = P\left[\begin{array}{cc} -s-1 & 0 \\ -s & 0 \\ \hdashline -s^2 & 1 \\ 0 & -1 \end{array}\right], \quad \text{and} \quad M(s) = \begin{bmatrix} -s^3 & s \\ 0 & -s \end{bmatrix}.$$

With $\varepsilon_1 = -\varepsilon_2 = 1$, $\alpha(s) = [\alpha_1 \quad 0]\begin{bmatrix} -s-1 \\ -s \end{bmatrix} = -\alpha_1 s - \alpha_1$, $\beta(s) = [0 \quad \beta_1]\begin{bmatrix} -s-1 \\ -s \end{bmatrix} = -\beta_2 s$, the matrix in (3.11) is $M_d(s) = \begin{bmatrix} \alpha_1 s + \alpha_1 + (s-\alpha_2)s^2 & (s-\alpha_2) \\ \beta_1 s & -(s-\beta_2) \end{bmatrix}$ which is equivalent to:

$$\begin{bmatrix} \alpha_1 + (\alpha_1 + \beta_1)s + (s-k)s^2 & 0 \\ \beta_1 s & -(s-k) \end{bmatrix},$$

with $\alpha_1 = \beta_2 = K$. Then, $h(s) = \alpha(s) + \beta(s) + (s-k)X(s) = [-\alpha_1 s - \alpha_1 - \beta_1 s + (s-K)s^2]$, and $\det M_d(s) = h(s)(s-K)$. Obviously, $X(s) = s^2$, $x_0 = 0$, and $\mu_1 = \mu_2 = 0$. The polynomial $h(s)$ is chosen equal to: $h(s) = (s+1)^3 = s^3 + 2s^2 + 3s + 1$, i.e. $h_0 = 3 = -(\lambda_1 + \lambda_2 + \lambda_3)$, with $\lambda_i = -1$ $(i = 1, 2, 3)$. The corresponding feedback matrix K_d is then

$$K_d = \begin{bmatrix} -1 & -3 & 0 & 0 \\ 0 & 0 & -2 & -3 \end{bmatrix}.$$

4. *Conclusions*

In this paper the problem of the pole placement with local static feedback has been considered. The problem has been examined only for two interconnected systems, and only for a special structure of system matrices (corresponding to the rank $[G] = 1$ case). Only $n-1$ poles can be arbitrarily placed, while the nth one must be real, and satisfy a certain relation. Obviously, this is the price to be paid for the assumption that two feedback coefficients are taken to be equal; an assumption which in fact transforms a nonlinear problem (as is the pole placement problem), into a linear one. It is believed that, in the general case (rank $[G] = 2$), all n eigenvalues can be arbitrarily placed with the n feedback coefficients. However, the formal proof of this statement remains open.

References

Anderson, B. D. O. and D. J. Clements (1979). Algebraic characterizations of fixed modes in decentralized control. *Automatica*, **15**, 703–711.

Calogiannis, J. and P. Fessas (1982). Fixed modes and local feedbacks in interconnected systems. *Int. J. of Control*, **37**, 175–182.

Chen, C T (1984). *Linear System Theory and Design*. Holt, Reinhard and Winston, New York.

Davison, E. and O. Ozguner (1983). Characterizations of decentralized fixed modes for interconnected systems. *Automatica*, **19**, 169–182.

Fessas, P. (1982). Stabilization with LLSVF: The state of the art and some open problems, *Fachberich Nr* 82-04 AIE/ETH; (1987) Stabilizability of two interconnected systems with local state-vector feedback: *Int. J. of Control*, **46**, 2075–2086; (1988) A generalization of some structural results to interconnected systems. *Int. J. of Control*, **43**, 1169–1186.

Kailath, T. (1980). *Linear Systems* Prentice Hall, Englewood Cliffs, NJ.

Wang, S. and E. Davison (1974). On the stabilization of the decentralized control systems. *IEEE on AC*, **18**, 473–478.

Wolovich, T. (1974). *Linear Multivariable Systems*. Springer Verlag, Berlin.

Automatica, Vol. 29, No. 2, pp. 469–474, 1993
Printed in Great Britain.

Brief Paper

Robust Eigenstructure Assignment via Dynamical Compensators*†

GUANG-REN DUAN‡

Key Words—Control system design; linear systems; eigenvalue assignment; sensitivity analysis; parameter optimization.

Abstract—Based on a proposed complete parametric approach for eigenstructure assignment in multivariable linear systems via dynamical compensators, insightful parametrizations of the closed loop eigenvalues sensitivities to the perturbed elements in the open loop system matrices are obtained, and an effective algorithm for eigenvalue assignment with minimum sensitivity in multivariable linear systems via dynamical compensators is then proposed. The algorithm does not contain 'going back' procedures, and allows the closed loop eigenvalues to be conveniently optimized within desired regions. A numerical example demonstrates its effect, simplicity and numerical property.

1. *Introduction*

EIGENVALUE ASSIGNMENT with minimum sensitivity in multivariable linear systems is an important problem in the field of robust control, and has now attracted much attention. Goursharankar and Ramar (1976), Cavin and Bhattacharyya (1983), Kautsky *et al.* (1985), Sun (1987), Chu *et al* (1984), Chen (1988), Byers (1989) and Duan (1992a) have considered the problem of minimizing eigenvalue sensitivity to model parameter variations in all the elements of the closed loop state matrix, while Goursharankar and Ramar (1976), Haiming Qiu and Gourishankar (1984), Song-Jiao Shi and Yue-Yun Wang (1985), Mielke *et al.* (1985) and Owens and O'Reilly (1989) have considered the problem of minimizing eigenvalue sensitivity to model parameter variations in some, but not all, of the elements of the open loop system matrices. Due to the fact that most of the practical systems often possess special structures and, the parameter perturbations are usually specially structured, the problem of minimizing eigenvalue sensitivity to model parameter variations in part of the elements of the open loop system matrices is more preferable than that of minimizing eigenvalue sensitivity to model parameter variation in all the elements of the closed loop state matrix in certain sense.

Existing solutions to the problem of eigenvalue assignment with minimum sensitivity in multivariable linear systems at present are mostly restricted to the case of state feedback (Goursharankar and Ramar, 1976; Cavin and Bhattacharyya, 1983; Kautsky *et al.*, 1985; Sun, 1987; Haiming Qiu and Gourishankar, 1984; Song-Jiao Shi and Yue-Yun Wang, 1985; Miekle *et al.*, 1985; Owens and O'Reilly, 1989) and output feedback (Chu *et al.*, 1984; Chen, 1988; Duan, 1992a) and are subject to the following limitations.

(1) Most of the existing solutions, except Owens and O'Reilly (1989) and Duan (1992a) are based on such eigenstructure assignment results with which only implicit relationship between the system matrices, the closed loop eigenspectrum and the design freedom can be established, and thus the robustness indexes as well as the finally converted optimization problems can not be arranged in forms explicitly expressed by the design free parameters to be optimized. This consequently results in algorithms that contain "going back" procedures. Such procedures not only add more computational burden to complement of the algorithm, but also may lead solutions to the problem far from optimal.

(2) Most of the existing solutions except Owens and O'Reilly (1989) and Duan (1992a) are, not at least in theory, proven to have utilized all the freedom existed in the design process. Therefore, no conclusion can be drawn about their optimality.

(3) Most of the existing solutions (except Duan 1992a) do not include the closed loop eigenvalues as optimizing parameters. While as a matter of fact, specific performance of a system allows the closed loop poles to be located in a certain region, but not necessarily at a few specified points. Since closed loop eigenvalues often appear in the robustness indexes with higher nonlinearity, proper small changes in the closed loop eigenvalues may significantly improve the robustness of the designed system (see our numerical example).

In this paper, we examine the problem of eigenvalue assignment with minimum sensitivity in multivariable linear systems via dynamical compensators. Since the closed loop system state matrix of a dynamical compensator controlled system is strongly structured, we restrict our attention to minimize eigenvalue sensitivity to model parameter variations in part of the elements of the open loop system matrices. We solve the problem based on a presented parametric eigenstructure assignment approach. For eigenstructure assignment in multivariable linear systems using dynamical compensators, Sambandan and Chandrasekharan (1981) and Han (1989) have developed non-parametric approaches, and Hippe and O'Reilly (1987) has presented a parametric approach for the case that the closed loop eigenvalues are distinct and are different from the open loop ones, and still the last one must be real. In this paper, another parametric approach to eigenstructure asignment via dynamical compensators is proposed based on an output feedback eigenstructure assignment result in Duan (1992a) (also refer to Duan *et al.*, 1991) which overcomes the few drawbacks in Hippe and O'Reilly (1987) (see Remark 3.2) and establishes the complete parametric representations for the closed loop eigenvectors and the compensator coefficient matrices with respect to the closed loop poles and four groups of partially free parameter vectors. Utilizing this parametric eigenstructure assignment result, and with the help of the Hellman–Feynman Theorem (Stephen, 1988) insightful parametrizations of the closed loop eigenvalue sensitivities to the perturbed elements in the open loop system matrices are obtained, and an effective algorithm for insensitive eigenvalue assignment in multivariable linear

* Received 2 October 1991; revised 7 April 1992; revised 23 July 1992; received in final form 23 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor T. Başar under the direction of Editor H. Kwakernaak.

† This paper has not been presented at any IFAC meetings.

‡ P.O. Box 329, Department of Control Engineering, Harbin Institute of Technology, Harbin 150006, P. R. China.

systems via dynamical compensators is then proposed. Due to the simplicity, explicity and the completeness of the proposed eigenstructure assignment approach, the obtained algorithm possesses several properties: (1) It is simple, numerically stable and requires less computational work. (2) It does not contain "going back" procedures since the optimization problem contained in the algorithm is in explicit forms with respect to the optimized parameters. (3) Its optimality is completely determined by the optimality of the solution to the optimization problem. (4) It gives systems with better robustness since closed loop eigenvalues can be conveniently included in the optimizing parameters and optimized within some desired specific regions. A numerical example demonstrates the above advantages.

2. *Problem formulation*

Consider the following linear system

$$\dot{x} = [A + \Delta A]x + [B + \Delta B]u, \quad y = [C + \Delta C]x \tag{1a}$$

where $x \in R^n$, $u \in R^r$, $y \in R^m$ $(r, m \leq n)$ are respectively the state vector, the input vector and output vector; A, B, C are known real matrices of appropriate dimensions and B, C are of full rank. ΔA, ΔB and ΔC are system parameter perturbations in the following form

$$\Delta A = \sum_{i=1}^{l} A_i \varepsilon_i, \quad \Delta B = \sum_{i=1}^{l} B_i \varepsilon_i, \quad \Delta C = \sum_{i=1}^{l} C_i \varepsilon_i \tag{1b}$$

here A_i, B_i and C_i, $i = 1, 2, \ldots, l$ are known real matrices of proper dimensions, ε_i, $i = 1, 2, \ldots, l$ are small perturbation variables.

When the following dynamical compensator

$$\dot{z} = K_{22}z + K_{21}y, \quad u = K_{11}y + K_{12}z \tag{2}$$

is applied to system (1), the closed loop system is obtained in the following form

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix} = [A_c + \Delta A_c]\begin{bmatrix} x \\ z \end{bmatrix} \tag{3}$$

with

$$A_c = \begin{bmatrix} A + BK_{11}C & BK_{12} \\ K_{21}C & K_{22} \end{bmatrix} \tag{4a}$$

$$\Delta A_c = \begin{bmatrix} \Delta A + BK_{11}\Delta C + \Delta BK_{11}C + \Delta BK_{11}\Delta C & \Delta BK_{12} \\ K_{21}\Delta C & 0 \end{bmatrix} \tag{4b}$$

and $z \in R^p$, K_{ij}, $i, j = 1, 2$ are real matrices of proper dimensions.

Noting the fact that nondefective matrices possess better robustness than defective ones (see Kautsky *et al.*, 1985) we describe the robust control problem for system (1) as follows.

Problem RC. Given system (1) and a proper region Ω in the left complex half plane, find real matrices K_{ij}, $i, j = 1, 2$ of proper dimensions such that the following conditions are met:

(1) Matrix A_c is nondefective, and has self conjugate eigenvalues all located in region Ω.

(2) The eigenvalues of matrix $A_c + \Delta A_c$ at $\varepsilon_i = 0$, $i = 1, 2, \ldots, l$ are as insensitive as possible to small variations in ε_i, $i = 1, 2, \ldots, l$.

Remark 2.1. The region Ω in Problem RC represents the requirement on the stability and performance of the closed loop system (3)–(4).

3. *Preliminary result*

Let $[A, B]$ be controllable, and $[A, C]$ be observable, then there hold the following right coprime factorizations

$$(sI - A)^{-1}B = N(s)D^{-1}(s) \tag{5}$$

$$(sI - A^T)^{-1}C^T = H(s)L^{-1}(s) \tag{6}$$

where $N(s) \in R^{n \times r}$, $D(s) \in R^{r \times r}$, $H(s) \in R^{n \times m}$ and $L(s) \in R^{m \times m}$ are all polynomial matrices, and $N(s)$ and $D(s)$, $H(s)$ and $L(s)$ are both right coprime.

Theorem 1. Let $[A, B]$ be controllable, $[A, C]$ be observable, $N(s)$, $D(s)$, $H(s)$ and $L(s)$ be matrix polynomials satisfying the right factorizations (5) and (6), and s_i, $i = 1, 2, \ldots, n + p$ be a group of self conjugate complex numbers (not necessarily distinct).

(1) There exist real matrices K_{ij}, $i, j = 1, 2$ of proper dimensions and matrices T, $V \in C^{(n+p)\times(n+p)}$ such that

$$A_c = V\Lambda T^T, \quad T^TV = I \tag{7}$$

with

$$\Lambda = \mathrm{diag}\,[s_1, s_2, \ldots, s_{n+p}] \tag{8}$$

if and only if there exist parameter vectors $f_{0i} \in C^r$, $g_{0i} \in C^m$, $f_{1i} \in C^p$, $g_{1i} \in C^p$, $i = 1, 2, \ldots, n + p$ satisfying the following constraints

(C1) $f_{ji} = \bar{f}_{jl}$, $g_{ji} = \bar{g}_{jl}$ if $s_i = \bar{s}_l$, $j = 1, 2$

(C2) $g_{0i}^T H^T(s_i)N(s_j)f_{0j} + g_{1i}^T f_{1j} = \delta_{ij}$, $i, j = 1, 2, \ldots, n + p$

where δ_{ij} is the Kronecker function.

(2) When constraints (C1) and (C2) are met, all the matrices T, V satisfying (7)–(8) are given by

$$V = \begin{bmatrix} V_0 \\ F_1 \end{bmatrix}, \quad T = \begin{bmatrix} T_0 \\ G_1 \end{bmatrix} \tag{9}$$

with

$$V_0 = [v_{01}\ v_{02}\ \cdots\ v_{0(n+p)}], \quad v_{0i} = N(s_i)f_{0i} \tag{10a}$$

$$T_0 = [t_{01}\ t_{02}\ \cdots\ t_{0(n+p)}], \quad t_{0i} = H(s_i)g_{0i} \tag{10b}$$

$$F_1 = [f_{11}\ f_{12}\ \cdots\ f_{1(n+p)}] \tag{11a}$$

$$G_1 = [g_{11}\ g_{12}\ \cdots\ g_{1(n+p)}] \tag{11b}$$

While the corresponding real matrices K_{ij}, $i, j = 1, 2$ are either given by

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = W[\Phi \quad (I - \Phi CV_0)\Psi] \tag{12}$$

with

$$\Psi = F_1^T(F_1F_1^T)^{-1}, \quad \Phi = \Gamma(CV_0\Gamma)^{-1} \tag{13a}$$

$$\Gamma = (I - \Psi F_1)(CV_0)^T \tag{13b}$$

$$W^T = [W_0^T \quad \Lambda F_1^T] \tag{13c}$$

$$W_0 = [w_{01}\ w_{02}\ \cdots\ w_{0(n+p)}], \quad w_{0i} = D(s_i)f_{0i} \tag{13d}$$

or by

$$\begin{bmatrix} K_{12} \\ K_{22} \end{bmatrix} = \begin{bmatrix} \Phi \\ \Psi(I - B^TT_0\Phi) \end{bmatrix}^T Z^T \tag{14}$$

with

$$\Psi = (G_1G_1^T)^{-1}G_1, \quad \Phi = (\Gamma T_0^TB)^{-1}\Gamma \tag{15a}$$

$$\Gamma = B^TT_0(I - G_1^T\Psi) \tag{15b}$$

$$Z^T = [Z_0^T \quad \Lambda G_1^T] \tag{15c}$$

$$Z_0 = [z_{01}\ z_{02}\ \cdots\ z_{0(n+p)}], \quad z_{0i} = L(s_i)g_{0i} \tag{15d}$$

Remark 3.1. The above Theorem 1 is a generalization of the output feedback eigenstructure assignment result in Duan (1992a) (Lemma 1). It can be proven by observing the fact that the design of a dynamical compensator of order p for system (1a) can be converted into the design of an output feedback controller for an extended order system with the following coefficient matrices (refer to Fletcher, 1980)

$$A_e = \mathrm{Diag}\,[A \quad 0_p], \quad B_e = \mathrm{Diag}\,[B \quad I_p], \quad C_e = \mathrm{Diag}\,[C \quad I_p] \tag{16}$$

and applying the Lemma 1 in Duan (1992a) and using the well known Matrix Inverse Lemma. It is notable that this result can be further generalized using the results in Duan *et al.* (1991) into the case where the closed loop matrix A_c possesses a general Jordan form.

Remark 3.2. This result clearly reveals all the degree of freedom in the dynamical compensator controlled nondefective systems, and gives the closed loop eigenvectors and the dynamical compensator matrices in clear neat forms which can be conveniently realized by computers. It differs with

that of Hippe and O'Reilly (1987) in several aspects. (a) It eliminates the strict conditions on the closed loop eigenvalues required in Hippe and O'Reilly (1987). (b) The approach of Hippe and O'Reilly (1987) requires the computation of inverses of $(n+p)$ number matrices of order $(n+p)$ and a matrix of order $(m+p)$. While our approach requires only some simple matrix elementary transformations (see Remark 3.3) and computation of inverses of two matrices of orders p and m (or r), respectively. (c) The approach of Hippe and O'Reilly (1987) requires some 'try and test' procedures while ours does not, when p is large enough, (say $p > n - m - r + 1$ (refer to Remark 3.4)).

Remark 3.3. A key step concerning the implementation of the above Theorem 1 is to obtain the polynomial matrices $N(s)$ and $D(s)$ satisfying (5). One of the simple and effective ways to achieve this is to find a pair of unimodular matrices $P(s)$ and $Q(s)$ satisfying the following equation:

$$P(s)[A - sI \quad B]Q(s) = [0 \quad I], \tag{17}$$

and then choose $N(s)$ and $D(s)$ as follows:

$$\begin{bmatrix} N(s) \\ D(s) \end{bmatrix} = Q_1(s), \quad Q(s) = [Q_1(s) \quad Q_2(s)]. \tag{18}$$

Subject to the controllability of $[A, B]$ equation (17) can be always realized simply by applying a series of elementary matrix transformations to the matrix $[A - sI \; B]$ (refer to Duan *et al.*, 1991; Duan, 1992a, b, c)

Remark 3.4. The issue of minimal order compensators design for eigenvalue assignment in linear systems, though studied by many authors (Kimura, 1975, Rosenbrock and Hayton, 1978), has never been completely solved. Observing the fact that a dynamical compensator with increased order, but not the minimal, provides additional useful degrees of design freedom, we may choose the dynamical order p according to any conservative result in the literature. It is notable that constraints (C1) and (C2) are met at least for the case of distinct closed loop eigenvalues when the compensator order p is large enough (say $p > n - m - r + 1$) to allow arbitrary assignment of the closed loop eigenvalues of the compensator controlled system.

4. *Solution to Problem RC*

Observing that the matrix A_c possessing eigenstructure (7)–(8) is nondefective. Then a main task left for solution to Problem RC is to establish the closed loop eigenvalue sensitivities to variables ε_i, $i = 1, 2, \ldots, l$ in the open loop system matrices. To achieve this purpose, we first present the following lemmas.

Lemma 1 (Hellman–Feynman Theorem (Stephen, 1988)). Let $M = (m_{ij}) \in R^{n \times n}$ be a nondefective matrix with eigenvalues λ_i, $i = 1, 2, \ldots, n$, and y_i and x_i be, respectively the left and the right eigenvectors of matrix M associated with eigenvalue s_i, $i = 1, 2, \ldots, n$. Then

$$\frac{\partial \lambda_k}{\partial m_{ij}} = \left(y_k^T \frac{\partial M}{\partial m_{ij}} x_k\right) \tag{19}$$

Lemma 2. Let $[A \; B]$ be controllable, $[A \; C]$ be observable, and constraints (C1) and (C2) be met. Then there hold

$$KC'V = W, \quad T^T B'K = Z^T, \tag{20}$$

where matrices B', C' are given by (16), while matrices V, T, K, W, Z are given as in Theorem 1.

Proof. Noting that similar relations hold for eigenstructure assignment via output feedback (Duan *et al.*, 1991; Kwon and Youn, 1987) then the conclusion follows recalling the fact mentioned in Remark 3.1.

With Lemma 1 and Lemma 2, we are now able to prove the following theorem.

Theorem 2. Subject to the assumptions and condition 1 of Theorem 3.1, the eigenvalue sensitivities of system (3)–(4) to variations ε_i, $i = 1, 2, \ldots, l$ are given as follows:

$$\bar{s}_{ij} = \frac{\partial s_i}{\partial \varepsilon_j} = d_{ij} + e_{ij}, \tag{21}$$

where

$$d_{ij} = g_{0i}^T[H^T(s_i)A_jN(s_i) + H^T(s_i)B_jD(s_i) + L^T(s_i)C_jN(s_i)]f_{0i}, \tag{22}$$

$$e_{ij} = g_{0i}^T H^T(s_i) \sum_{p=1}^{l} \varepsilon_p (B_jK_{11}C_p + B_pK_{11}C_j)N(s_i)f_{0i}, \tag{23}$$

$$i = 1, 2, \ldots, n+p, \quad j = 1, 2, \ldots, l.$$

Proof. Define

$$\Delta A' = \mathrm{Diag}[\Delta A \quad 0_p], \quad A'_i = \mathrm{Diag}[A_i \quad 0_p], \tag{24a}$$

$$\Delta B' = \mathrm{Diag}[\Delta B \quad 0_p], \quad B'_i = \mathrm{Diag}[B_i \quad 0_p], \tag{24b}$$

$$\Delta C' = \mathrm{Diag}[\Delta C \quad 0_p], \quad C'_i = \mathrm{Diag}[C_i \quad 0_p], \tag{24c}$$

then there holds

$$\Delta A' = \sum_{k=1}^{l} A'_k \varepsilon_k, \quad \Delta B' = \sum_{k=1}^{l} B'_k \varepsilon_k, \quad \Delta C' = \sum_{k=1}^{l} C'_k \varepsilon_k, \tag{25}$$

and

$$\begin{aligned} \Delta A_c &= \Delta A' + B'K\,\Delta C' + \Delta B'KC' + \Delta B'K\,\Delta C' \\ &= \sum_{k=1}^{l} (A'_k + B'KC'_k + B'_kKC')\varepsilon_k \\ &\quad + \sum_{k=1}^{l} \sum_{j=1}^{l} (B'_jKC'_k + B'_kKC'_j)\varepsilon_k \varepsilon_j. \end{aligned} \tag{26}$$

Therefore by applying Lemma 1 and using (26), we have

$$\bar{s}_{ij} = t_i^T \frac{\partial \Delta A_c}{\partial \varepsilon_j} v_i = d_{ij} + e_{ij},$$

with

$$d_{ij} = t_i^T(A'_j + B'KC'_j + B'_jKC')v_i,$$

$$e_{ij} = t_i^T \sum_{k=1}^{l} \varepsilon_k (B'_jKC'_k + B'_kKC'_j)v_i,$$

$$i = 1, 2, \ldots, n+p, \quad j = 1, 2, \ldots, l.$$

In view of Lemma 2, d_{ij} can be further converted into the following form:

$$d_{ij} = t_i^T A'_j v_i + t_i^T B'_j w_i + z^T C'_j v_i.$$

Finally using equation (24) and applying Theorem 1, we can convert d_{ij} and e_{ij} into the form of (22) and (23), respectively.

It follows from Theorem 1 and Theorem 2 that the key step in solving Problem RC is to perform the following optimization problem:

$$\begin{aligned} &\mathrm{Min} \quad F(f_{ji}, g_{ji}, s_i, j = 0, 1, i = 1, 2, \ldots, n+p) \quad (27) \\ &\mathrm{s.t.} \quad s_i \in \Omega, \quad i = 1, 2, \ldots, n+p. \\ &\qquad \text{Constraint } (C_1) \text{ and } (C_2) \end{aligned}$$

where

$$F(f_{ji}, g_{ji}, s_i, j = 0, 1, i = 1, 2, \ldots, n+p) = \sum_{i=1}^{n+p} \sum_{j=1}^{l} \alpha_{ij} d_{ij}^2, \tag{28}$$

with $\alpha_{ij} > 0$, $i = 1, 2, \ldots, n+p$; $j = 1, 2, \ldots, l$ being proper weighting factors. Here we have only considered the part of d_{ij} since it can be easily reasoned that d_{ij} is the dominant part in $\bar{s}_{ij}$ when ε_i, $i = 1, 2, \ldots, l$ are small.

Denote the real eigenvalue s_i of A_c by σ_i, and the corresponding parameters f_i and g_i by ξ_i and η_i, respectively; denote a pair of complex eigenvalues s_i and s_j of A_c by $s_i = \bar{s}_j = \sigma_i + \sigma_j i$, and the corresponding parameters by $f_{ji} = \bar{f}_{jl} = \xi_{ji} + \xi_{jl} i$, $g_{ji} = \bar{g}_{jl} = \eta_{ji} + \eta_{jl} i$, $j = 0, 1$, where σ_i, ξ_{ji} and η_{ji}, $j = 0, 1$, are real. Then constraint (C1) automatically holds and index (28) can be turned into the following form:

$$F = F(\xi_{ji}, \eta_{ji}, \sigma_i, j = 0, 1, i = 1, 2, \ldots, n+p).$$

Further by specially taking the convex field Ω as the sum of a series of square regions in the complex plane, the optimization problem (27) can be arranged into the following form:

$$\begin{aligned} &\text{Min } F(\zeta_{ji}, \eta_{ji}, \sigma_i, j = 0, 1, i = 1, 2, \ldots, n+p), \quad (29)\\ &\quad \text{s.t. constraint (C2)}\\ &\quad a_i \leq \sigma_i \leq b_i, \quad i = 1, 2, \ldots, n+p. \end{aligned}$$

Based on the above deduction and analysis, an algorithm for solution to Problem RC can be given as follows:

Algorithm RC

(1) Solve the right coprime factorizations (5) and (6).
(2) Solve the parametric expressions for matrices T_0, V_0, W_0 and Z_0 according to (10), (13d) and (15d).
(3) Give the values of a_i and b_i, $i = 1, 2, \ldots, n+p$ according to the stability and dynamical response characteristic requirements of the closed loop system.
(4) Establish the explicit expression of index F with respect to the optimizing parameters.
(5) Solve the optimization problem (29) by applying proper optimization algorithm.
(6) Calculate matrices T, V and K according to (9)–(11) and (12)–(13) or (14)–(15) based on the parameters obtained in Step (5).

Remark 4.1. The above algorithm has the following properties.
(a) It gives closed loop systems with better robustness (see the numerical results in the next section) since closed loop eigenvalues can be conveniently included into the optimizing parameters and optimized within desired regions.
(b) It possesses good numerical property since a well-conditioned solution to Step (1) can always be obtained due the lack of uniqueness of the elementary transformations which conduct matrix $[A - sI \;\; B]$ into $[0 \;\; I]$ (refer to the remarks in Section 4 of Duan (1992b)).
(c) It is simple since it does not contain 'going back' procedures and can be easily carried out in alphabetical order; it requires less computational work since its implementation involves only some matrix elementary calculations, the inverses of two matrices of reduced orders and the solution of an independent optimization problem.
(d) Its optimality is completely determined by the optimality of solution to the optimization problem in Step (4) due to the completeness of the eigenstructure assignment result.

Remark 4.2. The key step in our algorithm is the solution of the optimization problem (29) which directly determines the optimality of the solution to Problem RC. Unfortunately, it is hard to derive general definite conclusions about the optimality of the solution to problem (29) even when the closed loop eigenvalues are previously assigned to some specified points, since the problem is a general nonlinear programming with both the index and the constraints multilinear in variables and generally possessing no convexity. In practical applications, this optimization problem may be solved using any standard optimization algorithm. But before implementing this optimization it is important to simplify this optimization problem through the following two ways. (1) Properly fix an element in each f_i or h_i since there exists lack of uniqueness in eigenvectors of matrices (also refer to O'Reilly and Fahmy, 1985). (2) Solve as many as possible variables from constraint (C_2). These procedures will significantly reduce the number of optimizing parameters and the number of constraints in the optimization problem, and hence make the optimization much easier and provides solutions with higher precision.

5. *Example*

Consider a controllable and observable linear system with the following parameters.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When extended to four dimensions as in (16), this system becomes (A', B', C') which has been considered by Fletcher (1980), Chu *et al.* (1984), Chen (1988) and Duan (1992a). By the method given in Remark 3.3, we can easily obtain

$$N(s) = \begin{bmatrix} -s \\ -s^2 \\ 1 \end{bmatrix} \qquad D(s) = -s^3 + s^2 + s,$$

$$H(s) = \begin{bmatrix} s-1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad L(s) = \begin{bmatrix} s^2 - s - 1 & 1 \\ 0 & s \end{bmatrix}.$$

In the following, we consider the design of the first order dynamical compensator for this system. Restrict the closed loop poles s_i, $i = 1$–4 to be real, and let

$$f_{0i} = 1, \quad f_{1i} = x_i, \quad g_{0i}^T = [y_{i1} \;\; y_{i2}], \quad g_{1i} = y_{i3}, \quad i = 1\text{–}4,$$

then the right closed loop eigenvectors are given as

$$v_i^T = [-s_i \;\; -s_i^2 \;\; 1 \;\; x_i], \quad i = 1\text{–}4, \tag{30}$$

and constraint (C2) can be written in the following form:

$$s_j(s_i + s_j - 1)y_{i1} - y_{i2} - x_j y_{i3} = -\delta_{ij} \quad i, j = 1\text{–}4. \tag{31}$$

Define

$$s_{ijk} = (s_k - s_j)(s_i + s_j + s_k - 1),$$

$$s_{ij} = s_{iji}, \quad x_{ij} = x_i - x_j, \quad i, j, k = 1\text{–}4,$$

then the following insightful fact is valid (refer to Duan 1992a).

Fact. Let s_i, $i = 1$–4 be distinct and negative real, x_i, y_{ij}, $i = 1$–4, $j = 1$–3 be parameters satisfying the equation in (31). Then
(1) $s_{ij} \neq 0$, $s_{ijl} \neq 0$, $s_{ij}x_{il} - s_{il}x_{ij} \neq 0$ hold for i, j, l 1–4, $i \neq l \neq j$.
(2) Parameters x_i, y_{ij}, $i = 1$–4, $j = 1$–3 can be explicitly given through the following formulae.

$$x_2 = \frac{s_{324}x_1 + s_{321}x_4}{\ldots}, \quad x_3 = \frac{s_{234}x_1 + s_{213}x_4}{\ldots}$$

$$s_{12}x_{13} - s_{13}x_{12}$$

$$s_{21} - s_{24} \qquad x_{24}y_{23} - 1$$

$$- s_{24}x_{21}$$

$$s_{34} - s_{31} \qquad - 1$$

$$s_{34}x_{31} - s_{31}x_{34} \qquad s_{31}$$

$$y_{43} - \frac{s_{43} - s_{42}}{s_{43}x_{42} - s_{42}x_{43}} \qquad y_{41} \qquad \frac{x_{42}y_{43} - 1}{s_{42}},$$

$$y_{i2} = -s_i(2s_i - 1)y_{i1} + x_i y_i \quad 1, \quad i = 1 \;\; 4,$$

where x_1 and x_4 are two nonzero free real parameters.

It follows from the above fact that the optimization problem in Step (5) of our Algorithm RC for this example system has only six completely free parameters (including the closed loop poles) to be optimized.

Eleven solutions to eigenvalue assignment in this example system via the first order dynamical compensators are listed in Table 1 (where k_{ij} represents the element of the ith row and jth column in the constructed matrix K). Solutions $K_0 \sim K_5$ and K_8 are existing ones, while solutions K_6, K_7 and $K_9 \sim K_{10}$ are obtained by the presented approach under perturbations in the form of (1b) with

$$A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$B_3 = \qquad C_4 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} \tag{32a}$$

$$A_3 = A_4 = 0; \quad B_1 = B_2 = B_4 = 0; \quad C_1 = C_2 = C_3 = 0. \tag{32b}$$

TABLE 1. DYNAMICAL COMPENSATOR MATRICES

Solutions	k_{11}	k_{12}	k_{13}	k_{21}	k_{22}	k_{23}
K_0 (Fletcher (1980))	−47	34	10	49	−35	−11
K_1 (Chu *et al.* (1984))	−46.65	41.39	13.48	36.32	−31.69	−10.92
K_2 (Chu *et al.* (1984))	−47.00	41.70	13.63	36.52	−31.90	−11.00
K_3 (Chen (1988))	−46.9999	27.1833	−17 2069	−28.0806	−15.9828	−11.0000
K_4 (Chen (1988))	−46.9996	−10.4883	−24.3246	−18.3150	−5.72964	−10.9999
K_5 (Duan (1992a))	−47.00537	12.39453	23.88391	19.61450	−4.703613	−11.00049
K_6	−46.99937	8.847533	23.16630	20.06524	−3.165118	−10.99989
K_7	−47.00006	8.797701	5.108595	90.98519	−14.24602	−11.00000
K_8 (Duan (1992a))	−64.34406	22 44226	34.43155	23.33524	−7.737335	−13.51834
K_9	−58.75030	17.56029	22.24512	31.70565	−8.947386	−13.00002
K_{10}	−58.74984	17.57862	13.49337	52.27109	−14.76805	−12.99995

TABLE 2. CLOSED LOOP EIGENVALUES AND EIGENVECTORS

Solutions	s_1	s_2	s_3	s_4	x_1	x_2	x_3	x_4
K_0	−1	−2	−3	−4	1.4	7	14	23
K_1	−1.051842	−2.164131	−2.766958	−3.937068	0.659992	5.357691	8.439294	15.93948
K_2	−0.995089	−1.981681	−3.053435	−3.969795	0.443847	4.487640	10.018350	16.08444
K_3	−0.999940	−2.001396	−2.997297	−4.001368	−1.209604	−4.469316	−8.520009	−13.77098
K_4	−1.000517	−1.997385	−3.003993	−3.998005	−2.405559	−4.699992	−7.597358	−11.27596
K_5	−1.000740	−1.994811	3.010174	−3.994765	1.492578	3.822434	6.800663	10.51307
K_6	−0.9998724	−1.999867	−3.000609	−3.999352	1.689753	4.106956	7.130991	11.01099
K_7	−1.000958	−1.99304	−3 011227	−3.994784	7.683374	18.55133	32.51201	49.85145
K_8	−1.000079	−1.729420	−2.288860	−6.499980	1.246160	2.766928	6.746092	20.50924
K_9	−0.9999192	−1.499925	−3.000284	−6.499893	1.896369	3.35737	8.61836	30.32894
K_{10}	−0.9981721	−1.504096	−2.997073	−6.50061	3.116829	5.554411	14.18514	50.00907

Solution K_0 was obtained without consideration of robustness by a pole assignment approach, while solutions K_1–K_5 and K_8 were obtained by minimizing the condition numbers of the closed loop eigenvector matrices. Solutions K_0 – K_7 were all obtained on the condition that the closed loop eigenvalues are previously assigned to $s_1 = -1$, $s_2 = -2$, $s_3 = -3$, $s_4 = -4$, while solutions K_8 – K_{10} were obtained by optimizing closed loop eigenvalues within the following region.

$$-6.5 \le s_4 \le -4.5, \quad -4.5 \le s_3 \le -2.5, \quad -2.5 \le s_2 \le -1.5, \quad -1.5 \le s_1 \le -1.$$

The practical closed loop eigenvalues and eigenvectors corresponding to these solutions are given in Table 2 (it follows from (30) that parameters s_i and x_i completely determine the closed loop eigenvector v_i). Some robustness measures associated with these solutions are given in Table 3 (where $K_2(V) = \|V\|_2 \|V^{-1}\|_2$, $c = [c_1, c_2, \ldots, c_{n+p}]$, $c_i = \|t_i\| \|y_i\|$, $i = 1, 2, \ldots, n+p$). Table 4 shows the shifted closed loop eigenvavalues of the system under structural perturbation (1b) and (32) with $\varepsilon_1 = \varepsilon_2 = 0.05$, $\varepsilon_3 = \varepsilon_4 = 0.01$.

From Table 3 and Table 4, we can see the following points.

(a) Solutions with smaller $\|c\|_2$ and $K_2(V)$ values usually have smaller F values, but solutions with smaller F values do not necessarily possess small $\|c\|_2$ and $K_2(V)$ values.
(b) The imaginary parts of the shifted closed loop eigenvalues generally decreases and the real parts generally get closer to nominal values as the index F gets smaller.
(c) Inclusion of the closed loop eigenvalues into the optimizing parameters significantly improves the robustness of the closed loop system.

We have found through numerical computation that index $K_2(V)$ is more sensitive to the closed loop eigenvalues and eigenvectors than the other indexes appeared in Table 3, especially when its true value is large. Very small changes in the closed loop eigenvalues and/or eigenvectors generally give relatively small error in indexes F, $\|c\|_2$ and $\|K\|_F$, but may lead index $K_2(V)$ far away from its nominal value. Probably because of this reason, Chu *et al.* (1985) and Chen (1988) have all derived, for their solution, incorrect values of the measure $K_2(V)$ while giving the correct ones of measures

TABLE 3. ROBUSTNESS COMPARISON OF EXISTING SOLUTIONS

Solutions	F	$\|c\|_2$	$K_2(V)$	$\|K\|_F$
K_0	179.6541	454.21350	1016.8220	84.92349
K_1	255.7211	571.62100	1454.6020	80.70756
K_2	195.2416	428.45190	930.1136	81.27722
K_3	173.2308	333.61600	752.6219	66.40007
K_4	174.6446	346.41980	803.3182	58.30882
K_5	167.3580	300.87540	680.8795	58.83395
K_6	166.2896	301.16710	686.7354	57.94417
K_7	166.9077	705.20000	1515 8720	104.47370
K_8	48.9879	86.94968	557.6660	81.34181
K_9	40.59199	81 35593	693.3435	73.51208
K_{10}	40.62913	100.30980	770.5225	84.03561

TABLE 4. SHIFTED CLOSED LOOP EIGENVALUES UNDER SYSTEM PERTURBATIONS

Solutions	$s_{1,2}$	
K_0	$-1.1718552 \pm 0.2796065j$	$-3.828196 \pm 1.1704669j$
K_1	$-1.1708877 \pm 0.3763229j$	$-3.789154 \pm 1.2370240j$
K_2	$-1.1380780 \pm 0.3386860j$	$-3.8619442 \pm 1.2308790j$
K_3	$-1.2014933 \pm 0.1993400j$	$-3.7986069 \pm 1.1193070j$
K_4	-0.8852484 -2.0108064	$-3.552054 \pm 0.6713790j$
K_5	-1.0295902 -1.5206610	$-3.725676 \pm 0.9925040j$
K_6	-0.9958176 -1.5968800	$-3.703596 \pm 0.9529098j$
K_7	-0.9967350 -1.5944849	$-3.704390 \pm 0.9542037j$
K_8	-1.0229320 -1.4578643	-3.9413924 -6.0962797
K_9	-1.1052814 -1.1749888	-3.5209724 -6.1988390
K_{10}	-1.0949966 -1.1867652	-3.5184949 -6.1997573

$\|c\|_2$ and $\|K\|_F$. We have given much more precise closed loop eigenvalues and eigenvectors in Table 2 for all these solutions. In fact, it can be verified that with the values given in this table, $\det(A + BKC - s_iI)$ and $\|(A + BKC - s_iI)v_i\|$ are all in 10^{-50}, while the eigenvalues given in Chu *et al.* (1984) and Chen (1988), $\det(A + BKC - s_iI)$, $i = 1$–4 have only reached 10^{-1} and 10^{-20}, respectively. Therefore, our values given in Table 3 are much more accurate. Moreover, it can be seen from whichever sense that solution K_1 is not a good solution at all.

6. *Conclusion*

This paper has presented a simple, effective, complete parametric approach for eigenstructure assignment in multivariable linear systems via dynamical compensators. Using this approach, unified explicit and complete parametric representations of the compensator coefficient matrices and the closed loop eigenvector matrices can be established with respect to the closed loop poles and four groups of partially free parameter vectors which clearly reveal the design freedom in the problem. By utilizing this eigenstructure assignment result, the problem of robust dynamical compensator design, in the sense that the closed loop eigenvalues are as insensitive as possible to small variations in part of the elements of the open loop system matrices, is successfully treated. And an efficient algorithm is presented which allows closed loop poles to be optimized within desired regions, and is shown, either by analysis or computational results, to be simple, require less computational work and possess good numerical property and better optimality.

Acknowledgement—The author is grateful to the reviewers for their helpful comments and suggestions.

References

Byers, R. and S. G. Nash (1989). Approaches to robust pole assignment. *Int. J. Control*, **49**, 97–117.

Cavin, R. K. and S. P. Bhattacharyya (1983) Robust and well-conditioned eigenstructure assignment via Sylvesters equation. *Opt. Control Appl. and Methods*, **4**, 205–212.

Chen, C. H. (1988). On a numerical method for robust pole assignment by output feedback. *J. Comp. Math.* (in Chinese), **1**, 59–67.

Chu, E. K. W., N. K. Nichols and J. Kautsky (1984). Robust pole assignment by output feedback. *Proc. 4th IMA Conf. on Control Theory*, pp. 137–145.

Duan, G. R. (1992a). A simple algorithm for robust pole assignment in linear output feedback. *IEE Proc., Part D* **139**, 465–469.

Duan, G. R. (1992b). Solution to matrix equation $AV + EVF = BW$ and eigenstructure assignment for descriptor systems. *Automatica*, **28**, 639–643.

Duan, G. R. (1992c). Solution to matrix equation $AV + VF = BW$ and their application to eigenstructure assignment in linear systems. *IEEE Trans Aut. Control*, **37** (to appear).

Duan, G. R., G. Y. Wu and W. H. Huang (1991). Eigenstructure assignment for time-varying linear systems. *Sciences in China (Series A, English ed.)*, **34**, 246–256.

Fletcher, L. (1980). An intermediate algorithm for pole placement by output feedback in linear multivariable control systems. *Int. J. Control*, **31**, 1121–1136.

Gourishankar, V. and K. Ramar (1976). Pole assignment with minimum eigenvalue sensitivity to plant parameter variations. *Int J. Control*, **23**, 493–504.

Haiming Qiu and V Gourishankar (1984). Design of optimal feedback controllers for minimum eigenvalue sensitivity. *Opt. Control App. and Methods*, **5**, 309–317.

Han, Z. Z (1989). Eigenstructure assignment using dynamical compensator. *Int J Control*, **49**, 233–245

Hippe, P. and J. O'Reilly (1987). Parametric compensator design. *Int. J. Control*, **45**, 1455–1468.

Kautsky, J., N. K. Nichols and P. Van Dooren (1985) Robust pole assignment in linear state feedback *Int. J. Control*, **41**, 1129–1155.

Kimura, H. (1975). Pole assignment by gain output feedback. *IEEE Trans Aut. Control*, **AC-20**, 509–516

Kwon, B H and M. J. Youn (1987). Eigenvalue-generalized eigenvector assignment by output feedback. *IEEE Trans Aut Control*, **AC-32**, 417–421

Mieckle, R. R., P. I. Carraway and M. Marefat (1985). Interactive software for spectral assignment. *Proc. 24th CDC*. Ft. Lauderdale, FL, U.S.A., pp. 468–474.

O'Reilly, J. and M. M. Fahmy (1985). The minimum number of degrees of freedom in state feedback control. *Int. J. Control*, **41**, 749–768.

Owens, T. J. and J. O'Reilly (1989). Parametric state-feedback control for arbitrary eigenvalue assignment with minimum sensitivity. *IEEE Proc Part D: Control Theory and Applications*, **136**, 307–313.

Rosenbrock, H. H. and G. E. Hayton (1978). The general problem of pole assignment. *Int. J. Control*, **27**, 837–852.

Sambandan, A. and P. C. Chandrasekharan (1981). Eigenvector assignment by output feedback. *Int. J. Control*, **34**, 1143–1152.

Shi, S. J. and Y. Y. Wang (1985). Eigenstructure selection for rubust controller design. *Proc. IASTED. Int. Symp. Modelling and Simulation*, Lugano, Switzerland, pp. 289–292.

Stephen, B H. (1988). The generalized eigenproblem, pole-zero computation. *Proc. of the IEEE*, **76**, 103–120

Sun, J. G. (1987). On numerical methods for robust pole assignment in control system design. *J Comp. Math.*, **5**, 119–134.

Automatica, Vol. 29, No. 2, pp. 475–478, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Robust Disturbance Decoupling Problem for Parameter Dependent Families of Linear Systems*†

G. CONTE‡ and A. M. PERDON§

Key Words—Linear systems; multivariable control systems, control systems synthesis; robust control; disturbance rejection, state-space methods; state feedback.

Abstract—We consider the problem of disturbance decoupling by means of a single state feedback for a whole family of systems depending on a parameter, p. We give a geometric necessary and sufficient condition for the existence of solution to such problem under the hypothesis that the systems of the family are generically injective and the dependence on p is polynomial or, respectively, the set of possible values for p is finite.

1. *Introduction*

THE PROBLEM OF synthesizing a state feedback law which decouples the output of a system with respect to a nonmeasurable disturbance, or Disturbance Decoupling Problem (DDP), is a relevant one in control theory. In the case of linear systems, a complete solution to this problem can be given in geometric terms using the notion of controlled invariant subspaces (Basile and Marro, 1969; Wonham, 1985). Suitable generalizations of the geometric approach provide a solution to the DDP in several other contexts, concerning for instance nonlinear systems (Isidori *et al.*, 1981) linear periodic systems (Grasselli and Longhi, 1986), two-dimensional systems (Conte and Perdon, 1988), infinite dimensional systems (Curtain, 1986).

In this note, our aim is to study the DDP for linear systems in presence of uncertainty. More precisely we will consider a family of linear systems $\Sigma_d(p)$ described by the equations

$$\begin{cases} \dot{x}(t) = A(p)(t) + B(p)u(t) + D(p)w(t) \\ y(t) = C(p)x(t), \end{cases} \tag{1}$$

where $p = (p_1, \ldots, p_s) \in \Omega \subset R^s$ is a parameter whose presence models an uncertainty on the system coefficients, u is the control variable and w is a disturbance. The goal of the design problem is to find a static state feedback law $u = Fx$, not depending on p, such that for every p in Ω the output of the compensated system $\Sigma_{dF}(p)$ does not depend on the disturbance w. This problem, that we call the Robust Disturbance Decoupling Problem (RDDP), was first considered in Bhattacharyya (1983) in the case in which the matrices defining $\Sigma_d(p)$ depend linearly on the components p_is of p. The results obtained in Bhattacharyya (1983) consist essentially of a sufficient geometric condition for the existence of solutions and of a procedure for testing it. However, being equivalent to the existence of a single subspace having the same invariance property with respect to all the systems of the family, the sufficient condition of Bhattacharyya (1983) is far from being necessary. Subsequently, a special case, in which the matrices defining the systems of the family take values separately in a convex set, has been investigated in Ghosh (1985). In that paper it is assumed that $p = (p_1, p_2, p_3, p_4) \in \Omega = \{x = (x_1, x_2, x_3, x_4)$ such that $0 \le x_i \le 1\} \subset R^4$, $A(p) = A(p_1) = p_1 A_0 + (1 - p_1)A_1$, $B(p) = B(p_2) = p_2 B_0 + (1 - p_2)B_1$, $C(p) = C(p_3) = p_3 C_0 + (1 - p_3)C_1$, $D(p) = D(p_4) = p_4 D_0 + (1 - p_4)D_1$ and necessary and sufficient conditions for the existence of solutions to the RDDP are found.

Here, we strongly improve the results of Bhattacharyya (1983) by providing a necessary and sufficient condition for the existence of solutions to the Robust Disturbance Decoupling Problem in the case in which the matrices defining $\Sigma_d(p)$ depend polynomially on the components p_is of p, Ω is an open subset of R^s and generically (that is for all p except those belonging to the set of zeros of some polynomial) $\Sigma_d(p)$ is a left invertible system. Note that the last one, when $\Sigma_d(p)$ depends polynomially on the components of p, is a relatively mild assumption. Our results represent also an improvement of those of Ghosh (1985) since the family $\Sigma_d(p)$ we consider is more general than that in Ghosh (1985).

From a technical point of view, we start by considering the case (probably of more practical interest) in which Ω is a finite set, then we reduce the general case to that one. The tools we employ are those of the so-called geometric approach (see Wonham, 1985). In particular, in proving our results we make use of the notion of self-bounded control invariant subspace introduced and studied in Basile and Marro (1982). The intersection of two self-bounded controlled invariant subspaces has been shown (Basile and Marro, 1982) to be a controlled invariant subspace and this implies, in particular, the existence of the smallest controlled invariant subspace containing a given subspace. By means of this object, we can state a necessary and sufficient condition for the existence of solutions to the RDDP that, assuming left invertibility for the systems of the family $\Sigma_d(p)$, can be practically checked by means of classic geometric algorithms.

The paper is organized as follows. In Section 2, we state the problem and we describe some preliminary results on the set of the controlled invariant subspaces of a system. In particular, relying on the results of Basile and Marro (1982) on self-bounded controlled invariant subspaces, we show the existence of minimal elements of such set and recall a procedure, based on geometric algorithms, for constructing them. In Section 3, we provide our main result and we discuss an example.

Through the paper, we assume that the reader is familiar with the basic concepts of the geometric approach and, in particular, with the notions of controlled invariant subspace,

* Received 29 May 1991; revised 8 May 1992; received in final form 17 July 1992. This paper was recommended for publication in revised form by Associate Editor R. V. Patel under the direction of Editor H. Kwakernaak.

† A preliminary version of this paper has been presented at the First IFAC Symposium on Design Methods of Control Systems, Zurich, Switzerland, September 1991.

‡ Dipartimento di Elettronica e Automatica, Università di Ancona, Via Brecce Bianche, 60131 Ancona, Italy.

§ Dipartimento di Matematica, Università di Ancona, via Brecce Bianche, 60131 Ancona, Italy.

conditionally invariant subspace and with the related algorithms, as they are described, for instance, in Basile and Marro (1969) and Wonham (1985).

2. *Preliminaries and statement of the problem*

Let us consider the family of linear control system $\Sigma_d(p)$ described by

$$\begin{cases} \dot{x}(t) = A(p)x(t) + B(p)u(t) + D(p)w(t) \\ y(t) = C(p)x(t) \end{cases} \tag{2}$$

where $x \in X = R^n$ is the state, $u \in U = R^m$ is the control, $w \in W = R^h$ is a disturbance, $y \in Y = R^q$ is the output, $p = (p_1, \ldots, p_s) \in \Omega \subset R^s$ is a parameter and $(A(p), B(p), C(p), D(p))$ are matrices of suitable fixed dimensions whose entries depend on p.

The control problem concerning the pair $(\Sigma_d(p), \Omega)$ we want to consider is defined as follows:

2.1. *Definition.* Given the family of systems $\Sigma_d(p)$ described by (1) and the value set Ω for the parameter p, the Robust Disturbance Decoupling Problem (RDDP) concerning the pair $(\Sigma_d(p), \Omega)$ consists in finding a static state feedback law $u = Fx$, not depending on p, such that for every p in Ω the output of the compensated system $\Sigma_{dF}(p)$ does not depend on the disturbance w.

2.2. *Remark.* One may prefer to refer to the above design problem as to a "simultaneous" or "parameter independent" disturbance decoupling problem, using, as done in Basile and Marro (1987) the adjective robust to denote the case in which the parameter p is supposed to vary with time and the output of the compensated system is required to be independent on the disturbance with respect to such variation. No matter what terminology one adopts, it should be clear that our problem is equivalent to solving, by the same feedback, a family of disturbance decoupling problems, one for each value of p in Ω, while the one considered in Basile and Marro (1987) is essentially a disturbance decoupling problem for a time-varying system, whose acceptable solutions are time-varying feedbacks.

It is quite easy to derive from Basile and Marro (1969) or Wonham (1985) the following result.

2.3. *Proposition.* The RDDP concerning the pair $(\Sigma_d(p), \Omega)$ is solvable if and only if there exists a family $\{V(p)\}_{p \in \Omega}$ of subspaces of X such that:

(i) $\operatorname{Im} D(p) \subset V(p) \subset \operatorname{Ker} C(p)$ for all $p \in \Omega$;
(ii) there exist a linear map $F: X \to U$ such that $(A(p) + B(p)F)V(p) \subset V(p)$ for all $p \in \Omega$.

By generalizing, in a straightforward way, the classic terminology of the geometric approach, a family of subspaces $\{V(p)\}_{p \in \Omega}$ for which the Condition 2.3 (ii) holds will be called a feedback-type family of controlled invariant subspaces, and any map $F: X \leftarrow U$ such that $(A(p) + B(p)F)V(p) \subset V(p)$ for all $p \in \Omega$ will be called a friend of the family. Clearly, the difficulty in dealing with the RDDP is in testing the Conditions 2.3 (i) and (ii). In order to describe a procedure for doing so, at least for a meaningfully large class of families $\Sigma_d(p)$, we need to introduce some auxiliary results about the set of the controlled invariant subspaces of a linear system. To fix the notation, let us consider the linear system

$$\Sigma = \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t), \end{cases} \tag{3}$$

with $x \in X = R^n$. Following the terminology introduced in Basile and Marro (1982), a controlled invariant subspace V for Σ is called self-bounded if it contains $(V^* \cap \operatorname{Im} B)$. The basic properties of the self-bounded controlled invariant subspaces are studied in Basile and Marro (1982); here we recall in particular the following results.

2.4. *Proposition.* Let V^* denote the maximum controlled invariant subspace for Σ contained in $\operatorname{Ker} C$, and let $K \subset V^*$ be a subspace for which the condition $(V^* \cap \operatorname{Im} B) \subset K$ holds. Then, the set of all the controlled invariant subspaces for Σ containing K and contained in $\operatorname{Ker} C$ has a minimal element.

Proof. Any controlled invariant subspace V containing K and contained in $\operatorname{Ker} C$ is self-bounded, hence the conclusion follows from Basile and Marro (1982).

Assuming that $(V^* \cap \operatorname{Im} B) \subset K$ holds for a given subspace K of V^*, let us denote by $V_*(K)$ the minimal element in the set of all the controlled invariant subspaces for Σ containing K. The following result of Basile and Marro (1982) relates $V_*(K)$ to other geometric objects.

2.5. *Proposition* (Basile and Marro, 1982). Given Σ and a subspace $K \subset V^*$ with $(V^* \cap \operatorname{Im} B) \subset K$, let $S^*(K + \operatorname{Im} B)$ denote the minimum conditionally invariant subspace for Σ containing $(K + \operatorname{Im} B)$. Then, $V_*(K)$ coincides with the intersection $V^* \cap S^*(K + \operatorname{Im} B)$.

The above proposition points out, in particular, that by using the well-known geometric algorithms, which provide, respectively V^* and, with a suitable initialization, S - $(K + \operatorname{Im} B)$, (see Basile and Marro, 1969) one can construct $V_*(K)$.

If Σ is left invertible, since $V^* \cap \operatorname{Im} B = \{0\}$, the condition $(V^* \cap \operatorname{Im} B) \subset K$ and the conclusions of Proposition 2.4 and Proposition 2.5 hold for any subspace K of V^* (in particular all the controlled invariant subspaces of $\operatorname{Ker} C$ are self-bounded). Assuming left invertibility of Σ, one has in addition that B is injective and, hence, the following result.

2.6. *Proposition* (see also Schumacher, 1983). Let Σ be left invertible and let $V \subset V^*$ be a controlled invariant subspace for Σ. Then, we have the following.

(i) If $F: X \to U$ is a friend of V, F is also a friend of any controlled invariant subspace $V' \subset V$.
(ii) If $F: X \to U$ is a friend of V and $F': X \to U$ is a friend of $V' \subset V$, then $F_{|V'} = F'_{|V'}$

Proof (i) Let $(A + BF)V \subset V$ and $(A + BF')V' \subset V'$ hold for F and for some $F'. X \to U$ Then, for any $v' \in V' \subset V$, we have $Av' = v_1 + BFv' = v_2 + BF'v'$ with $v_1 \in V$ and $v_2 \in V' \subset V$. Since $V \cap \operatorname{Im} B = \{0\}$, this implies in particular that $v_1 = v_2 \in V'$ and hence $(A + BF)V' \subset V'$.
(ii) By the same argument as above, $BFv' = BF'v'$ and hence $Fv' = F'v'$ for any $v' \in V'$

3. *Main result*

Let us start by considering the RDDP concerning a pair $(\Sigma_d(p), \Omega_f)$, where $\Sigma_d(p)$ is defined by (1) and $\Omega_f = \{P_0, \ldots, P_r\} \subset R^s$ is a finite set.

3.1. *Proposition* Given $(\Sigma_d(p), \Omega_f)$ let us assume that

$$\Sigma(p) \cdot \cdot \begin{cases} \dot{x}(t) = A(p)x(t) + B(p)u(t) \\ y(t) = C(p)x(t) \end{cases}$$

(that is the undisturbed system) is left invertible for any $p \in \Omega_f$. Then, the RDDP concerning $(\Sigma_d(p), \Omega_f)$ is solvable if and only if $\operatorname{Im} D(p) \subset V^*(p)$ for all $p \in \Omega_f$ and $\{V_*(\operatorname{Im} D(p))\}_p \in \Omega_f$ is a feedback-type family of controlled invariant subspaces

Proof. The 'if' part is obvious. Conversely, let us assume that the RDDP at issue is solvable. By Proposition 2.3 there exists a feedback-type family of controlled invariant subspaces $\{V(p)\}_{p \in \Omega_f}$ such that $\operatorname{Im} D(p) \subset V(p)$ for all $p \in \Omega_f$. In particular, $\operatorname{Im} D(p)$ is contained in $V^*(p)$ and, hence, $V_*(\operatorname{Im} D(p))$ is well defined for all p in, in particular, $V_*(\operatorname{Im} D(p)) \subset V(p)$. Let F be a friend of $\{V(p)\}_{p \in \Omega_f}$; by Proposition 2.6 (i) F is also a friend of $V_*(\operatorname{Im} D(p))$ for any $p \in \Omega_f$ and, therefore $\{V_*(\operatorname{Im} D(p))\}_{p \in \Omega_f}$ is feedback-type family of controlled invariant subspaces

The relevance of the above result is due to the fact that the necessary and sufficient condition described can be practically checked. There is actually no difficulty in checking its first part, while its second part can be checked by using the following procedure.

3.2. *Procedure.* Step (0). Compute $V_*(\operatorname{Im} D(P_0))$ and a friend F_0 of it. Set $X_0 = V_*(\operatorname{Im} D(P_0))$ and $\mathbf{F}_0 = F_0$.
Step $(i + 1)$. Compute $V_*(\operatorname{Im} D(P_{i+1}))$ and a friend F_{i+1} of it. If $\mathbf{F}_{i|X_i \cap V_*(\operatorname{Im} D(P_{i+1}))} = F_{i+1|X_i \cap V_*(\operatorname{Im} D(P_{i+1}))}$, write $X = (X_i + V_*(\operatorname{Im} D(P_{i+1}))) \oplus X'_i$, set $X_{i+1} = X_i + V_*(\operatorname{Im} D(P_{i+1}))$ and

define $\mathbf{F}_{i+1}: X \to U$ by $\mathbf{F}_{i+1|X_i} = \mathbf{F}_{i|X_i}$, $\mathbf{F}_{i+1|V_*(\mathrm{Im}\,D(P_{i+1}))} = F_{i+1|V_*(\mathrm{Im}\,D(P_{i+1}))}$ and $\mathbf{F}_{i+1|X_i'} = \mathbf{F}_{i|X_i'}$. Otherwise, stop.

It is, in fact, easy to realize by Proposition 2.6(ii) that, if the above procedure stops because $\mathbf{F}_{i|X_i \cap V^*(\mathrm{Im}\,D(P_{i+1}))}$ is different from $F_{i+1|X_i \cap V_*(\mathrm{Im}\,D(P_{i+1}))}$, then $\{V_*(\mathrm{Im}\,D(p))\}_{p \in \Omega_f}$ is not a feedback-type family of controlled invariant subspaces. Otherwise, i.e. if the procedure can be performed up to Step (r) included, it eventually produces a friend $F = \mathbf{F}_r$ of $\{V_*(\mathrm{Im}\,D(p))\}_{p \in \Omega_f}$.

3.3. *Remark*. In general, very little can be said in geometric terms about the possibility of achieving the robust decoupling of the disturbance with stability, except—quite obviously—that a necessary condition is the solvability of the DDP with stability for any p (Wonham, 1985). Assuming that the above procedure yields a friend $\mathbf{F}_r$ of $\{V_*(\mathrm{Im}\,D(p))\}_{p \in \Omega_f}$ and letting X'' denote the subspace $(\bigcup_{p \in \Omega_f} V_*(\mathrm{Im}\,D(p)))$, any other friend coincides with $\mathbf{F}_r$ on X''. Then, a stronger necessary condition is that, for any $p \in \Omega_f$ the eigenvalues of the restriction of $(A(p) + B(p)\mathbf{F}_r)$ to the largest controlled invariant subspace of $\Sigma(p)$ contained in X'' have negative real part.

3.4. *Remark*. It is obvious from 3.2 that $\{V_*(\mathrm{Im}\,D(p))\}_{p \in \Omega_f}$ is a feedback type family if, in particular, each element in it has zero intersection with union of all the others. It follows then that the RDDP is generically solvable in the space of all families $\Sigma_d(p)$ verifying the conditions that for any p the corresponding DDP is solvable and that $\dim X \geq \Sigma_{p \in \Omega_f} \dim V_*(\mathrm{Im}\,D(p))$.

We can now remove the finiteness assumption on the value set for p. So, let us consider the RDDP concerning a pair $(\Sigma_d(p), \Omega)$, where $\Sigma_d(p)$ is defined by (2) and $\Omega \subset R^s$ is an open set.

3.5. *Proposition* Given $(\Sigma_d(p), \Omega)$ let us assume that

(i) the entries of the matrices $(A(p), B(p), C(p), D(p))$ defining $\Sigma_d(p)$ depend polynomially on the components of $p = (p_1, \ldots, p_s)$;

(ii) there exists a polynomial $P(p)$ such that

$$\Sigma(p) = \begin{cases} \dot{x}(t) = A(p)x(t) + B(p)u(t) \\ y(t) = C(p)x(t), \end{cases}$$

is left invertible for any $p \in \Omega \setminus \{p \in \Omega$ such that $P(p) = 0\}$.

Denoting by $\max \deg_{p_i} M(p)$ the maximum degree in p_i of the elements of a polynomial matrix $M(p)$, let r_i be an integer greater than or equal to

$$(\max \deg_{p_i} C(p) + \max \deg_{p_i} D(p)$$
$$+ (n-1) \max(\max \deg_{p_i} A(p), \max \deg_{p_i} (B(p)))$$

and chose arbitrary distinct values $a_{10}, \ldots, a_{1r_1}; a_{20}, \ldots, a_{2r_2}; \ldots; a_{s0}, \ldots, a_{sr_s} \in R$ such that the finite set $\Omega_f = \{P_{i_1 \ldots i_s} = (a_{1i_1}, \ldots, a_{si_s}) \in R^s$ for $i_j = 0, \ldots, r_j\}$ is contained in $\Omega \setminus \{p \in \Omega$ such that $P(p) = 0\}$. Then, the RDDP concerning $(\Sigma_d(p), \Omega)$ is solvable if and only if the RDDP concerning $(\Sigma_d(p), \Omega_f)$ is solvable. Moreover, any solution of one of the problems is a solution of the other.

Proof. The only 'if' part of the first group of statements is obvious, as well as the fact that any solution of the RDDP concerning $(\Sigma_d(p), \Omega)$ is a solution of the RDDP concerning $(\Sigma_d(p), \Omega_f)$. So, let us assume that the RDDP concerning $(\Sigma_d(p), \Omega_f)$ is solvable and let $F: X \to U$ be a solution This implies that the elements of the matrices $C(p)D(p)$, $C(p)(A(p) + B(p)F)D(p)$, $\ldots$, $C(p)(A(p) + B(p)F)^{n-1}D(p)$ are zero for all $p \in \Omega_f$. Since these elements are polynomials of degree lesser than or equal to r_i in each p_i, this implies (see the Appendix) that they are zero for all $p \in R^s$ and, in particular, for all $p \in \Omega$. Hence, the output of the compensated system $\Sigma_{dF}(p)$ does not depend on the disturbance w for any $p \in \Omega$ or, in other words, F is a solution of the RDDP concerning $(\Sigma_d(p), \Omega)$

As a result of the above proposition, the existence of solutions to the RDDP concerning $(\Sigma_d(p), \Omega)$ can be checked by means of the Procedure 3.2 with $\Omega_f = \{P_{i_1 \ldots i_s} = (a_{1i_1}, \ldots, a_{si_s}) \in R^s$ for $i_j = 0, \ldots, r_j\}$. Obviously this implies a large number of computations, but it also provides an algorithmic construction of a solution, if there are solutions.

3.6. *Example*. Let us consider the family of systems $\Sigma_d(p)$ described by (2) for

$$A(p) = \begin{pmatrix} 1 & 0 & p^2 \\ 0 & p+1 & 0 \\ p-1 & 1 & 0 \end{pmatrix} \quad B(p) = \big|$$

$$D(p) = \begin{pmatrix} p \\ -p^2 \\ 0 \end{pmatrix}, \quad C(p) = (0 \quad 0$$

and $p \in \Omega = R$. For $p = 0$, $\Sigma(0)$ is easily seen to be left invertible, hence $\Sigma(p)$ is generically left invertible. The dependence of $\Sigma_d(p)$ on p is not linear, nor is there a single controlled invariant subspace containing $\mathrm{Im}\,D(p)$ for all $p \in \Omega$ having a friend which does not depend on p. Hence, $\Sigma_d(p)$ does not verify the sufficient condition described in Bhattacharyya (1983) for the existence of solutions to the RDDP. Actually, the computation shows that $\mathrm{Ker}\,C$ is controlled invariant for all $p \in \Omega$, and hence $V^*(p) = \mathrm{Ker}\,C$ for all $p \in \Omega$, but $\{V^*(p)\}_{p \in \Omega}$ is not a feedback-type family of controlled invariant subspaces. Since $r_1 = 4$, we chose $\Omega_f = \{0\text{–}4\}$ and we compute $V_*(\mathrm{Im}\,D(p))$ for $p \in \Omega_f$ The computation yields

$$V_*(\mathrm{Im}\,D(0)) = \{0\},$$

$$V_*(\mathrm{Im}\,D(1)) = \mathrm{span}\left\{\begin{pmatrix} -1 \\ 0 \end{pmatrix}\right.$$

$$V_*(\mathrm{Im}\,D(2)) = \mathrm{span}\left\{\begin{pmatrix} -4 \\ 0 \end{pmatrix}\right.$$

$$V_*(\mathrm{Im}\,D(3)) = \mathrm{span}$$

$$V_*(\mathrm{Im}\,D(4)) = \mathrm{span}$$

and it is not difficult to verify that $F = (1 \quad 0 \quad 0): X \to U$ is a friend for $V_*(\mathrm{Im}\,D(p))$ for all $p \in \Omega_f$. Hence $\{V^*(p)\}_{p \in \Omega_f}$ is a feedback-type family of controlled invariant subspaces and, by Proposition 3.5, $F = (1 \quad 0 \quad 0)$ is a solution to the RDDP concerning $(\Sigma_d(p), R)$. In fact, as one can easily check, $C(p)(A(p) + B(p)F)^i D(p) = 0$ for all $p \in \Omega$ and all $i \geq 0$.

4 *Conclusion*

We have given a geometric necessary and sufficient condition for decoupling, by means of a single state feedback, the disturbance from the output of a parameter dependent family of linear systems. The main assumptions we made are that the systems in the family have coefficients which depend polynomially on the parameter and that they are generically left invertible The left invertibility is used for inferring the existence of minimal controlled invariant subspaces and for stating the problem in geometric terms. The assumption about polynomial dependence allows us to reduce the general case to the case of a finite family of systems and to compute practically a solution to the problem.

Acknowledgement—The authors wish to thank the unknown referees for valuable comments.

References

Basile, G and G. Marro, (1969). Controlled and conditioned invariant subspaces in linear system theory. *J. Optimization Theory and Application*, **3**.

Basile, G. and G. Marro (1982). Self-bounded controlled

invariant subspaces: a straightforward approach to constrained controllability. *J. of Optimization Theory and Applications*, **38**.

Basile, G. and G. Marro (1987). On the robust controlled invariant. *Systems & Control Letters*, **9**.

Bhattacharyya, S. (1983). Generalized controllability, (A, B)-invariant subspaces and parameter invariant control. *SIAM J. Alg. Disc. Meth.*, **4**.

Conte, G. and A. M. Perdon (1988). A geometric approach to the theory of 2D systems. *IEEE Trans. Aut. Control*, **AC-33**.

Curtain, R. (1988). Invariance concepts in infinite dimension. *SIAM J. of Control and Opt.*, **24**

Ghosh, B. K. (1985). A geometric approach to simultaneous system design: parameter insensitive disturbance decoupling by state and output feedback. *Proc. 7th Int. Symp. on Mathematical Theory of Networks and Systems*, Stockholm, Sweden.

Grasselli, O. and S. Longhi (1986). Disturbance localization with dead-beat control for linear periodic discrete-time systems. *Int. J. Control*, **44**.

Isidori, A., A Krener, C. Gori-Giorgi and C. Monaco (1981). Nonlinear decoupling via feedback: a differential geometric approach. *IEEE Trans. Aut. Control*, **AC-26**.

Schumacher, J. M. (1983). On a conjecture of Basile and Marro. *J. of Optimization Theory and Applications*, **41**.

Wonham, W. (1985). *Linear Multivariable Control. A Geometric Approach*. Springer-Verlag, NY.

Appendix

Proposition. Let $\Pi(p)$ be a polynomial in the components p_is of $p = (p_1, \ldots, p_s)$ with degree lesser than or equal to r_i in each p_i. Let $a_{10}, \ldots, a_{1r_1}$; $a_{20}, \ldots, a_{2r_2}; \ldots$; $a_{sr_s} \in R$ be arbitrary distinct values and consider the finite set $\Omega_f \subset R^s$ defined by $\Omega_f = \{P_{i_1 \ldots i_s} = (a_{1i_1}, \ldots, a_{si_s}) \in R^s$ for $i_j = 0, \ldots, r_j\}$. If $\Pi(P) = 0$ for all $P \in \Omega_f$, then $\Pi(p)$ is the identically zero polynomial, that is $\Pi(P) = 0$ for all $P \in R^s$.

Proof. We use induction on the number of variables. The statement is true for $s = 1$, since, in this case, $\Pi(p)$ is a polynomial of degree lesser than or equal to r_1 which vanishes in $r_1 + 1$ distinct points. Assume now that the statement is true for $s = k$ and let $p = (p_1, \ldots, p_{k+1})$. Write $\Pi(p)$ as a polynomial in p_{k+1} with coefficients $\Pi_j(\bar{p})$, the $\Pi_j(\bar{p})$ being polynomials in $\bar{p} = (p_1, \ldots, p_k)$ of degree lesser than or equal to r_i in each p_i. By evaluating the coefficients $\Pi_j(\bar{p})$ in $(a_{10}, a_{20}, \ldots, a_{k0})$, we get a polynomial $\Pi'(p_{k+1})$ in one variable, of degree lesser than or equal to r_{k+1} which vanishes in $r_{k+1} + 1$ distinct points. This implies that $\Pi'(p_{k+1})$ is the identically zero polynomial and, hence, that all the coefficients $\Pi_j(\bar{p})$ vanish in $(a_{10}, a_{20}, \ldots, a_{k0})$. By repeating the same argument, we get that each coefficient $\Pi_j(\bar{p})$ vanishes in all $P \in \Omega_f' = \{P_{i_1 \ldots i_k} = (a_{1i_1}, \ldots, a_{ki_k}) \in R^k$ for $i_j = 0, \ldots, r_j\}$. Then, by the induction hypothesis, all the $\Pi_j(\bar{p})$ are identically zero polynomials, and as a consequence $\Pi(p)$ is the identically zero polynomial.

Automatica, Vol. 29, No. 2, pp. 479–484, 1993
Printed in Great Britain

0005-1098/93 \$6.00 + 0.00

Brief Paper

Stability Robustness Characterization and Related Issues for Control Systems Design*†

YANLIN LI‡ and E. BRUCE LEE‡

Key Words—Robustness; optimization; linear systems; control system design.

Abstract—Two issues are addressed in this paper. First, we show that the H^∞ norm of a sensitivity function matrix for a stable multivariable closed loop system indicates its stability robustness in the sense of gain margin and phase margin. Second, we apply this result to the problem of synthesizing a robust stabilizing controller. The emphasis of the analysis is LQ feedback system design. The goal is to design an observer based controller such that the resulting closed loop system has the robustness property that an LQ state feedback system has. This is achievable if the sensitivity function of the observer based control system is very close to that of the state feedback system. The closeness is measured by the H^∞ norm and classical H^∞ optimization techniques are used for design.

1. *Introduction*

CONSIDER A FEEDBACK system with the configuration shown in Fig. 1. Here $K(s)$ is the transfer function of a stabilizing controller and $P(s)$ is the plant model. One fundamental concern in feedback control system design is stability robustness. Gain and phase margins are frequently used as indicators for stability robustness. In classical frequency domain SISO control systems synthesis, the robustness issue is easily handled by employing synthesis techniques based on Nyquist plots (or Bode or Nichols plots). In this paper, we use gain margin and phase margin to describe stability robustness.

For SISO cases, stability robustness in the sense of gain and phase margins can be described, by the nearest distance, say r_0, from the point $(-1, 0)$ in the complex plane to the Nyquist plot of the feedback system. Suppose that the closed loop system in Fig. 1 is stable and its Nyquist plot does not meet the circle with center at $(-1, 0)$ and radius r_0. In this case, it is known that the system has gain margin $(1/(1+r_0), 1/(1-r_0))$ and phase margin $\theta_0 = \pm\cos^{-1}(1-(r_0^2/2))$ (Lehtomaki *et al.* (1981)). Therefore gain and phase margins have an explicit relationship with the radius r_0 of such a circle which the Nyquist plot of a SISO system does not meet. We call r_0 the simultaneous stability margin. It can be easily shown that the nearest distance from the Nyquist plot to $(-1, 0)$ point is the inverse of the H^∞ norm of the system's sensitivity function. Hence, the H^∞ norm of a system's sensitivity function has an explicit relationship with gain and phase margin for a SISO feedback system.

The definition of gain and phase margin for an MIMO system means that simultaneous gain and phase variations in all the loops of a multivariable feedback system (Lehtomaki *et al.* (1981)). One motivation of this paper is to show that it is also true that H^∞ norm of a system's sensitivity function has an explicit relationship with gain and phase margin for MIMO systems.

A typical example revealing the relationship between the sensitivity function and stability robustness is LQ state feedback systems. If one uses LQR design technique in the SISO case, then the obtained feedback system has gain margin $(\frac{1}{2}, \infty)$ and phase margin $\theta_0 = \pm 60°$. This is because the H^∞ norm of the sensitivity function is less than or equal to one (use Kalman inequality), or the Nyquist plot does not pass through the unit circle, with radius $r_0 = 1$ and center at $(-1, 0)$. Thus a system with LQR controller has simultaneously stability margin $r_0 = 1$.

Since the robustness property of a feedback system can be described by its sensitivity function, one can design a control system with good robustness by putting some constraints on the sensitivity function. This provides us with a guide on how to design a feedback system with guaranteed stability margin (simultaneous gain and phase margins). Characterizing stability robustness by gain and phase margins is practically useful in the sense that if one knows how big the infinity norm of the sensitivity function is (by proper system design), then one has an idea of how much to adjust the loop gain to achieve a desired response without destabilizing the closed loop system.

In the last half of the paper, we investigate a well-known question: when can an observer based control system achieve the stability robustness property of an LQ state feedback control system? From the preceding discussion, one can conclude that this is achievable if the sensitivity function of the observer based control system is properly designed so that it is close to that of state feedback system in the sense of H^∞ norm. This leads to the idea of sensitivity recovery. It will be shown that sensitivity recovery implies loop transfer recovery. Thus we obtain a systematic way of doing LTR system design. An interesting result is that if the controlled system is minimum phase, then loop/sensitivity recovery can be achieved as well as desired, which is the same as the classical result. If the controlled system is non-minimum phase, then the recovery error can not be made arbitrarily small. However, the error can be minimized by using an H^∞ optimization approach, which is better that the classical result. This proposed recovery approach is more systematic than the classical approach (Doyle, 1981).

The notation to be used is standard: The infinity norm of $G(s)$ is defined by $\|G\|_\infty := \sup\{\bar{\sigma}(G(j\omega)) : \omega \in R\}$, where $\bar{\sigma}$ is the maximal singular value. The real part and imaginary part of G are denoted by $\mathrm{Re}(G)$ and $\mathrm{Im}(G)$, respectively. The determinant of G is denoted by $\det(G)$. Single-input/single-output and multi-input/multi-output are denoted by SISO and MIMO, respectively.

The paper is organized as follows. Stability robustness issues will be discussed in Section 2. In Section 3 we resolve some interesting issues posed in Section 2, and in particular propose a methodology to design an observer based controller that has good stability robustness, which leads to sensitivity and loop recovery. Illustrative examples are offered in Section 4.

* Received 28 February 1991; revised 14 January 1992; revised 14 July 1992; received in final form 23 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor H. Kimura under the direction of Editor H. Kwakernaak.

† This research was supported by the NSF under grants numbered DMS-87 22402 and DMS-90 02919.

‡ Department of Electrical Engineering, University of Minnesota, 200 Union St. Se, Minneapolis, MN 55455, U.S.A.

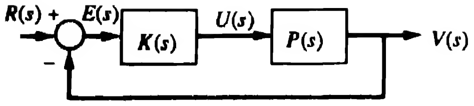


FIG. 1. Feedback systems configuration.

2. *Stability robustness characterization*

Definition 1. For a stable SISO feedback system as diagrammed in Fig. 1, the quantity r_0 with $0 < r_0 < 1$ is said to be a simultaneous stability margin if and only if

$$1 - re^{j\theta} + K(s)P(s) \neq 0 \quad \text{for all } 0 < \theta \leq 2\pi, \qquad 0 \leq r \leq r_0,\ \mathrm{Re}\,(s) \geq 0. \tag{1}$$

The above definition can be understood in the following way. Assume that a closed-loop system is stable. Then its Nyquist plot will not meet the point $(-1, 0)$. Simultaneous stability margin means that the Nyquist plot does not meet the point $(-1, 0)$ and further, it is kept a distance r_0 away from the $(-1, 0)$-point.

If the Nyquist plot is kept away from the $(-1, 0)$-point by at least r_0, then the system (SISO) will simultaneously possess the following gain margin and phase margin (Lehtomaki *et al.*, 1981):

$$\mathrm{GM} = \left(\frac{1}{1+r_0}, \frac{1}{1-r_0}\right),$$

$$\mathrm{PhM} = \pm\cos^{-1}\left(1 - \frac{r_0^2}{2}\right).$$

This means that if the loop transfer function $K(s)P(s)$ is perturbed to $K(s)P(s)\alpha e^{j\phi}$, then the closed loop system remains stable for all (ϕ, α) such that

$$-\cos^{-1}\left(1 - \frac{r_0^2}{2}\right) < \phi < +\cos^{-1}\left(1 - \frac{r_0^2}{2}\right),$$

and

$$\frac{1}{1+r_0} < \alpha < \frac{1}{1-r_0}.$$

Definition 2. For the stable MIMO feedback system as configured in Fig. 2 with m inputs and p outputs having $p \geq m$, the quantity r_0 with $0 < r_0 < 1$ is said to be a simultaneous stability margin in if and only if

$$|I + K(s)P(s) + URV| \neq 0 \quad \text{for } \mathrm{Re}\,(s) \geq 0, \tag{2}$$

where URV is the perturbation in polar form with $U^*U = I$, $VV^* = I$ and $\|R\|_\infty \leq r_0$.

Definition (2) means that in the multivariable case, a simultaneous stability margin of r_0 implies that if the loop transfer function $K(s)P(s)$ is perturbed to

$$K(s)P(s)\,\mathrm{diag}\,(l_1, l_2, \ldots, l_m),$$

(see Fig. 2), then closed-loop stability still remains, provided for all $i = 1, 2, \ldots, m$,

$$-\cos^{-1}\left(1 - \frac{r_0^2}{2}\right) < \phi_i < +\cos^{-1}\left(1 - \frac{r_0^2}{2}\right),$$

for $l_i = e^{j\phi_i}$ (pure phases);

$$\frac{1}{1+r_0} < l_i < \frac{1}{1-r_0}, \quad \text{for } l_i\text{s}$$

to be real numbers (pure gains).

This will be explained in more detail later.

Remark 1. In Definitions 1 and 2, we consider only the case $0 < r_0 < 1$. One can also consider the case when $r_0 > 1$, in somewhat the same way.

Remark 2. In Definition 2, dual results hold if $m \geq p$. The only difference is that the sensitivity function is $(I + P(s)K(s))^{-1}$ instead of $(I + K(s)P(s))^{-1}$, and the perturbation is considered at the output of the system.

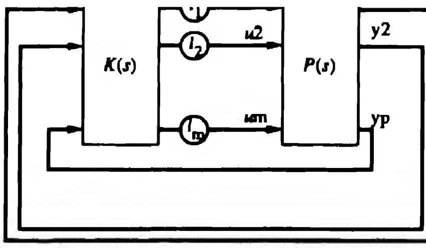


FIG. 2. MIMO feedback system configuration.

Theorem 2.1. Consider a SISO controlled system as configured in Fig. 1, $P(s)$ is the transfer function model of the controlled system and $K(s)$ is the transfer function of a stabilizing controller. This feedback control system has a simultaneous stability margin r_0 if and only if the sensitivity function $S(s) := [1 + K(s)P(s)]^{-1}$ satisfies

$$\|S(s)\|_\infty < \frac{1}{r_0}. \tag{3}$$

Proof. It is similar to that of Theorem 2.2.

Next we are going to present a result which bridges the gap between characterizing stability robustness of SISO systems and that of MIMO systems. This is done by Theorem 2.2 below. One feature of Theorem 2.1 and Theorem 2.2 is that one can characterize stability robustness of a closed loop system in a uniform framework with only the knowledge of nominal system model and controller (if only analysis is required, then only nominal plant model is necessary). We do not have to make assumptions on uncertainties that the control system may face. This is different from the usual approaches. Stability robustness adopted here for MIMO systems is characterized by gain margin and phase margin in each loop. This robustness characterization is more straightforward for engineers than those characterized by gap metric, additive or multiplicative uncertainties (Georgiou and Smith, 1990; Vidyasagar and Kimura, 1986) because engineers usually prefer to know how much gain change can be made to achieve a given goal without making the system unstable. It will be seen that this simple stability robustness result has important applications in H_∞/LQG/LTR design and stability robustness analysis.

Theorem 2.2. Consider linear MIMO feedback systems as suggested by Fig. 2 and let $K(s)$ be a stabilizing controller. Let URV be the loop perturbation in polar form with $U^*U = I$, $VV^* = I$ and $\|R\|_\infty < r_0$. Then system has a simultaneous stability margin r_0, i.e.

$$|I + K(s)P(s) + URV| \neq 0 \quad \text{for } \mathrm{Re}\,(s) \geq 0,$$

if and only if $\|S(s)\|_\infty < 1/r_0$, where $S(s) := [I + K(s)P(s)]^{-1}$ is the sensitivity function.

Proof.

$|I + K(s)P(s) + URV| \neq 0$ for $\mathrm{Re}\,(s) \geq 0$ if and only if

$|I + S(s)URV| \neq 0$ for $\mathrm{Re}\,(s) \geq 0$,

since $I + K(s)P(s)$ is stable by assumption.

Sufficiency. Assume $\|S(s)\|_\infty < 1/r_0$, then

$$\|S(s)URV\|_\infty \leq \|S(s)\|_\infty \sigma_{\max}(URV) < \left[\frac{1}{r_0}\right]|r_0| = 1. \tag{4}$$

Hence, $|I + K(s)P(s) + URV| \neq 0$, for $\mathrm{Re}\,(s) \geq 0$.

Necessity. We show this part by contradiction. Suppose that $|I + K(s)P(s) + URV| \neq 0$ for all $\mathrm{Re}\,(s) \geq 0$ does not imply $\|S(s)\|_\infty < 1/r_0$. Then $\|S(s)\|_\infty \geq 1/r_0$. Thus there exists a frequency ω_0 such that $\bar{\sigma}(S(j\omega_0)) = 1/r_0$, and $S(j\omega_0)$ can be decomposed as $S(j\omega_0) = U_0\Sigma V_0$, (singular value decomposition), where U_0 and V_0 are unitary and $\Sigma = \mathrm{diag}\,(1/r_0, \Sigma_1)$.

Now take $U = -V_0^*$ and $V = U_0^*$, then

$$|I + K(j\omega_0)P(j\omega_0) + URV| = |S(j\omega_0)^{-1} + URV|$$
$$= |V_0^*\Sigma^{-1}U_0^* - V_0^*RU_0^*|$$
$$= |V_0^*(\Sigma^{-1} - R)U_0^*| = 0,$$

since $\Sigma^{-1} - R = \mathrm{diag}(0, *)$, which contradicts the assumption. □

Corollary 2.3. If an MIMO feedback system has simultaneous stability margin r_0, then simultaneously in each loop of the feedback system of Fig. 2, there is a guaranteed gain margin (denoted GM) and phase margin (denoted PhM) given by

$$\mathrm{GM} = \frac{1}{1 \pm r_0}, \tag{5}$$

$$\mathrm{PhM} = \pm\cos^{-1}\left(1 - \frac{r_0}{2}\right) \tag{6}$$

The following lemma is cited with little change from Lehtomaki *et al.* (1981) which will be used in the proof of Corollary 2.3.

Lemma 2.4. For square matrices G and L, $\det(I + GL) \neq 0$ if

$$\sigma((I + G)^{-1})\sigma(L^{-1} - I) < 1 \tag{7}$$

Proof of Corollary 2.3 (cf. Lehtomaki *et al.* (1981)). Let $G(s)$ denote $K(s)P(s)$. Assume that $G(s)$ is perturbed to $G(s)L$, where

$$L = \mathrm{diag}(l_1, l_2, \ldots, l_m)$$

Suppose that one has found an appropriate controller such that $\|(I + G)^{-1}\|_\infty < 1/r_0$. Then by Lemma 2.4 $\det(I + GL) \neq 0$ for $\mathrm{Re}(s) > 0$ if

$$|(l_i^{-1} - 1)| < r_0 \tag{8}$$

for all i since (8) implies (7). To obtain gain margins let all l_is be real. From (7) one has

$$\frac{1}{1 + r_0} < l_i < \frac{1}{1 - r_0}$$

Similarly, let l_is be $e^{j\phi_i}$ where ϕ_i is real. From (7) one get the phase margins expression. □

Note $\mathrm{GM} = 1/(1 \pm r_0)$ is interpreted as that any gain α_i inserted in any one of the feedback loops in Fig. 2 satisfying $1/(1 + r_0) < \alpha_i < 1/(1 - r_0)$ will not destabilize the closed loop system. Similar interpretation holds for PhM $= \pm\cos^{-1}(1 - (r_0^2/2))$. Every feedback loop may allow a phase factor $e^{j\phi_i}$ to be inserted, provided $|\phi_i| < \cos^{-1}|(1 - (r_0/2))|$ so that closed-loop system remains stable.

3. *Stability margins for LQG control systems and robustness recovery*

Consider a MIMO feedback system with state space model

$$\frac{dx}{dt} = Ax + Bu \tag{9}$$

$$y = Cx \tag{10}$$

where $x \in R^n$, $B \in R^{n\times m}$, $C \in R^{p\times n}$ and $p \geq m$. Assume that (A, B, C) is a stabilizable and detectable triple. It is known that an observer based controller for this system has the form $F(sI - A + BF + LC)^{-1}L$, where $A - BF$ and $A - LC$ are both stability matrices, and $F = B'P$, $L = QC'$, with $P \geq 0$ and $Q \geq 0$ satisfying

$$A'P + PA - PBB'P + C'C = 0 \tag{11}$$

$$AQ + QA' - QC'CQ + BB' = 0 \tag{12}$$

respectively. Our goal is to present an idea for designing an observer based controller. We will show how to obtain an observer based control system which has guaranteed stability margin. And then we show how to design an observer such that the output feedback system recovers the stability robustness of an LQ state feedback system.

Fact. Let $P(s) = C(sI - A)^{-1}B$ be the transfer function model of the controlled system, and

$$K(s) = F(sI - A + BF + LC)^{-1}L,$$

be the compensator transfer function. Then

(i) $T(s) = [I + K(s)P(s)]^{-1}K(s)P(s) = F(sI - A + BF)^{-1} \times LC(sI - A + LC)^{-1}B$, and

(ii) $S(s) = [I + K(s)P(s)]^{-1} = \{I - F(sI - A + BF)^{-1}B\}\{I + F(sI - A + LC)^{-1}B\}$, where $S(s)$ is the sensitivity function and $T(s)$ is the complementary sensitivity function.

Proof. (i)

$$T(s) = [I + F(sI - A + BF + LC)^{-1}LC(sI - A)^{-1}B]^{-1}$$
$$\times F(sI - A + BF + LC)^{-1}LC(sI - A)^{-1}B$$
$$= F[I + (sI - A + BF + LC)^{-1}LC(sI - A)^{-1}BF]^{-1}$$
$$\times (sI - A + BF + LC)^{-1}LC(sI - A)^{-1}B$$
$$= F[sI - A + BF + LC(sI - A)^{-1}$$
$$\times (sI - A + BF)]^{-1}LC(sI - A)^{-1}B$$
$$= F(sI - A + BF)^{-1}\{I + LC(sI - A)^{-1}\}^{-1}$$
$$\times LC(sI - A)^{-1}B$$
$$= F(sI - A + BF)^{-1}LC\{I + (sI - A)^{-1}LC\}^{-1}(sI - A)^{-1}B$$
$$= F(sI - A + BF)^{-1}LC(sI - A + LC)^{-1}B$$

Fact (ii) can be easily verified by noticing $T(s) + S(s) = I$. □

It is known from Theorem 2.2 that stability robustness for a MIMO feedback system is completely characterized by the infinity norm of the sensitivity function. So we focus on sensitivity function analysis. If one uses an LQG controller, then the infinity norm of the sensitivity function is

$$\|S(s)\|_\infty = \|\{I - F(sI - A + BF)^{-1}B\}$$
$$\times \{I + F(sI - A + LC)^{-1}B\}\|_\infty, \tag{13}$$

and it can be computed by an existing method if F and L are known (Boyd *et al.*, 1989). The inverse of $\|S(s)\|_\infty$ is the simultaneous stability margin r_0. From (13) the infinity norm of the sensitivity function is bounded by

$$\|S(s)\|_\infty \leq \|\{I - F(sI - A + BF)^{-1}B\}\|_\infty$$
$$\times \|\{I + F(sI - A + LC)^{-1}B\}\|_\infty$$
$$\leq \|\{I + F(sI - A + LC)^{-1}B\}\|_\infty,$$

(using Kalman inequality). The next lemma is a special form of a result due to Mageirou and Ho (1977) which is useful in deriving the main result in this section.

Lemma 3.1. Let A, B, K and I (identity matrix) be matrices of compatible dimensions. Then the following statements are equivalent:

(i) A is a stability matrix and $\|\{I + K(sI - A)^{-1}B\}\|_\infty < 1/r_0$.

(ii) $r_0 < 1$ and there exists a positive definite symmetric matrix $X > 0$ such that

$$A'X + XA + (XB + K')(B'X + K)\left(\frac{1 - r_0^2}{r_0^2}\right) + I = 0$$

Theorem 3.1. If the following two Riccati equations

$$A'X + XA - XBB'X + C'C = 0, \tag{14}$$

$$AY + YA' - 2YC'CY + (YK' + B)$$
$$\times (KY + B')\left(\frac{1 - r_0^2}{r_0^2}\right) + I = 0 \tag{15}$$

have positive definite solutions X and Y, respectively, then the observer based controller $F(sI - A + BF + LC)^{-1}L$ with $F = B'X$ and $L = YC'$ will be a stabilizing controller and the obtained feedback system will have simultaneous stability margin r_0.

Proof. Assume that the system is controllabe and observable. If (14) has a positive definite solution then $A - BF$ is a

stability matrix and if (15) has a positive definite solution then the matrix $A - LC$ is a stability matrix. By separation principle, $F(sI - A + BF + LC)^{-1}L$ is a stabilizing controller. It is easy to see that (15) having a positive definite solution implies

$$\|I + F(sI - A + LC)^{-1}B\|_\infty < \frac{1}{r_0}. \quad (16)$$

Hence the sensitivity function of the observer based control system satisfies

$$\|S(s)\|_\infty \leq \|\{I + F(sI - A + LC)^{-1}B\}\|_\infty < \frac{1}{r_0}. \quad (17)$$

Thus, by Theorem 2.2, the obtained feedback system will have simultaneous stability margin r_0. □

Another observation can be made from (13). If F and L are chosen properly such that $F(sI - A + LC)^{-1}B$ is very 'small' then the sensitivity function of the observer based control system will approach the sensitivity function of the system with state feedback. Therefore, the robustness property of a state feedback system is recovered. This is the key idea for robustness recovery. We will use the infinity norm to measure the 'smallness'. Then H^∞ optimization technique can be used to design an output feedback system such that it recovers the stability robustness property of LQ state feedback systems.

Let us assume that the system (9), (10) satisfies the following four conditions.

(A1) (A, B) is stabilizable and (C, A) is detectable.
(A2) C and B are of full rank, respectively.
(A3) $m \leq p \leq n$.
(A4) The transfer function $C(sI - A)^{-1}B$ is left invertible and strictly minimum phase, i.e

$$\operatorname{rank}\begin{bmatrix} A - sI & B \\ C & 0 \end{bmatrix} = n + p,$$

for all Re $(s) \geq 0$.

The following result will be used in describing our observations concerning loop recovery and robustness recovery. This result is due to Mageirou and Ho (1977). Similar results with different settings can also be found in Petersen and Hollot (1988) and Zhou (1988).

Lemma 3.2. Assume conditions (A1)–(A4) hold, then for any $\delta > 0$ there exists a $q^* > 0$ such that the Riccati equation

$$AY + YA' + Q + q^2\frac{BB'}{\delta} - YC'CY + \frac{YF'FY}{q^2\delta} = 0, \quad (18)$$

has a positive-definite symmetric solution for all $q \geq q^*$, where Q is a given positive definite matrix. Furthermore, with L defined as $L := YC'$, the transfer function

$$G(s) := F(sI - A + LC)^{-1}B,$$

satisfies the bound

$$\|G(j\omega)\|_\infty \leq \delta.$$

Remark. Lemma 3.2 means that given any $\delta > 0$, Riccati equation (18) has a positive-definite solution provided $q > 0$ is chosen to be sufficiently large. It also means that if (A1)–(A4) are satisfied, then $\|G(j\omega)\|_\infty$ can be made arbitrarily small by choosing L properly. Denote by S_{sf} the sensitivity function of an LQ state feedback system, and denote by S_{of} the sensitivity function of a feedback system with observer based controller. Then, from the previous discussion, it follows that

$$S_{of}(s) = [I + K(s)P(s)]^{-1} = \{I - F(sI - A + BF)^{-1}B\} \times \{I + F(sI - A + LC)^{-1}B\}, \quad (19)$$

$$S_{sf}(s) = \{I - F(sI - A + BF)^{-1}B\}. \quad (20)$$

To make the stability robustness of an observer based control system recovered to be that of a state feedback system, one needs to choose L (observer gain) suitably; such that $A - LC$ is a stability matrix and $\|S_{obs} - S_{sf}\|_\infty$ is minimized. Note that

$$\begin{aligned}\|S_{obs} - S_{sf}\|_\infty &= \|\{I - F(sI - A + BF)^{-1}B\} \\ &\quad \times \{F(sI - A + LC)^{-1}B\}\|_\infty \\ &\leq \|I - F(sI - A + BF)^{-1}B\|_\infty \\ &\quad \times \|F(sI - A + LC)^{-1}B\|_\infty \\ &\leq \|F(sI - A + LC)^{-1}B\|_\infty. \end{aligned} \quad (21)$$

To minimize $\|S_{obs} - S_{sf}\|_\infty$ it is sufficient to minimize $\|F(sI - A + LC)^{-1}B\|_\infty$ by choosing L appropriately. This is a typical H infinity control task. From inequality (21) several observations can be made which are useful.

(A) If one can find a L such that $\|F(sI - A + LC)^{-1}B\|_\infty$ can be made as small as possible subject to $A - LC$ being a stability matrix, then the stability robustness property, which an LQ state feedback control system has, can be recovered as completely as possible.

(B) From Lemma 3.2, if the Assumptions (A1)–(A4) are satisfied, then there always exists L such that $\|F(sI - A + LC)^{-1}B\|_\infty$ is arbitrarily small with $A - LC$ a stability matrix.

(C) Let $G_{sf}(s)$ be the loop transfer function of a feedback system with LQ state feedback and $G_{of}(s)$ be the loop transfer function corresponding to the observer based control system. Suppose that (A1)–(A4) are satisfied and $G_{sf}(s)$ and $G_{of}(s)$ have no poles on the $j\omega$ axis, then S_{of} approaching S_{sf} in terms of the L_∞ norm implies that $G_{sf}(s)$ approaches $G_{of}(s)$ since

$$\begin{aligned}\|G_{obs} - G_{sf}\|_\infty &= \|[I + G_{obs}] - [I + G_{sf}]\|_\infty \\ &= \|(I + G_{obs})(S_{obs} - S_{sf})(I + G_{sf})\|_\infty \\ &\leq \|(I + G_{obs})\|_\infty \, \|(S_{obs} - S_{sf})\|_\infty \, \|(I + G_{sf})\|_\infty \\ &\leq \lambda \, \|(S_{obs} - S_{sf})\|_\infty, \end{aligned}$$

where λ is a constant. Thus robustness recovery implies loop recovery.

(D) If the plant is of non-minimum phase, then there does not exist an L such that $\|F(sI - A + LC)^{-1}B\|_\infty$ is made arbitrarily small with $A - LC$ a stability matrix. However one can iteratively solve for L such that $\|F(sI - A + LC)^{-1}B\|_\infty$ is minimized subject to $A - LC$ being a stability matrix. The minimum error bound between S_{of} and S_{sf} can be therefore obtained and the degree of loop transfer recovery can be characterized.

From the above observations, the following important facts are apparent.

(1) If the model of the plant is of minimum phase, then the stability robustness property of an LQ state feedback control system possesses can be achieved asymptotically by an observer based control system provided that $\|F(sI - A + LC)^{-1}B\|_\infty$ is made arbitrarily small, where F is the LQ state feedback gain matrix and L is the filter gain properly chosen

(2) If the model of the plant is of minimum phase, then loop transfer recovery can be achieved by an H_∞ optimization approach.

(3) If the model is of non-minimum phase, the observer based control system can never achieve the stability robustness property that a full LQ state feedback control system has, i.e the simultaneous stability margin is always less than one.

(4) If the model is of non-minimum phase, one can also use the LTR technique suggested here. The advantage of this approach is that it provides a way of minimizing the error of recovery, which is the key difference between our method and the existing ones. However, the lower error bound can never be zero. Its value depends highly on the structure of the non-minimal phase zeros.

We now introduce a new two-step LQG/LTR design procedure as suggested by the above.

Step (1). Design an LQ regulator by solving for the positive definite solution X of

$$A'X + XA + N'N - XBB'X/\rho = 0,$$

where N and ρ are free design parameters and (A, N) is

a detectable pair, with desirable loop transfer functions and $A - BF$ a stability matrix, where $F = B'X/\rho$.
Step (2). Design a sequence of filters by solving

$$AY + YA' + Q + q^2\frac{BB'}{\delta} - YC'CY + YK'KY$$

where δ is a small positive number, $Q > 0$ and q is used to help search for solutions, with $A - LC$ a stability matrix and $L = YC'$. When the plant is of minimal phase, δ can be chosen as small as necessary.

The compensator produced by these two steps has the following well-known form:

$$K(s) = F(sI - A + BF + LC)^{-1}L.$$

Remark 4.1. The LQG/LTR approach suggested here is essentially a combination of H^2 and H^∞ optimization designs. In Step (1) one solves an H^2 optimization task and in Step (2) one solves an H^∞ optimization task. If the design objective is H_∞ control, then in Step 1 one can choose F using H^∞ state feedback techniques.

Remark 4.2. The two-step LQG/LTR design procedure has its dual form, i.e. one can design the filter first via filter H^2 Riccati equation and then design a sequence of LQ regulators by H^∞ optimization.

Remark 4.3. This LTR procedure can also be used to recover a full state feedback H^∞ control loop transfer function.

4. *Illustrative examples*

Two examples are offered below to demonstrate how to design an observer based control system using the suggested approach such that the closed loop system has good stability robustness in the sense of stability margin. Example 1 deals with a minimum phase controlled system and Example 2 considers the case in which the controlled system is of non-minimum phase.

Example 1. Consider a controlled system as given by the transfer function

$$P(s) = \frac{s^2 + 12s + 20}{s^3 + 7.2s^2 + 13.4s + 2.4},$$

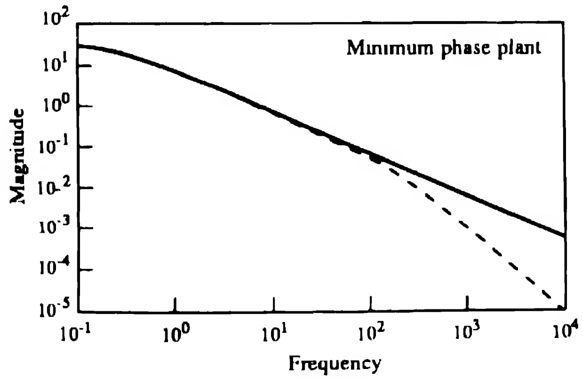


FIG. 3. Bode plot of target loop and recovery loop.

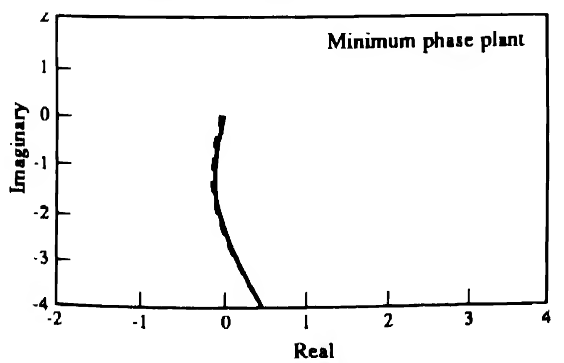


FIG. 4. Nyquist plot of systems with state feedback and output feedback.

which is of minimum phase. We want to design an observer based controller such that the feedback system is stable, has good stability robustness and satisfies given performance requirements. By 'good stability robustness', we mean that the Nyquist plot of the observer based feedback sytem will be kept away from the point $(-1, 0)$, or it almost does not go into the unit circle centered at the point $(-1, 0)$. By the 'performance requirements' we mean the target loop Bode plot is asymptotically recovered. By the two-step design procedure suggested in Section 4, the controller transfer function is found to be

$$F(sI - A + BF + LC)^{-1}L,$$

where (A, B, C) are obtained as controller form realization of $P(s)$; F and L are the state feedback gain and filter gain, respectively with

$$F = [6.2604, 54.6716, 87.0749],$$
$$L = [150.3813, 1.1590, 0.1230]'.$$

Figure 3 is the Bode plot which shows the target loop transfer function has been recovered. The solid line in Fig. 3 represents the Bode plot of the target loop with state feedback and dashed line is the Bode plot of loop transfer function with output feedback. Figure 4 is the Nyquist plot of the system with state feedback and observer based output feedback. It shows that the Nyquist plot with observer based output feedback (dashed line) only just goes into the unit circle centered at $(-1, 0)$ thus the observer based control system has good stability robustness.

Example 2. We now consider the case in which the controlled system is of non-minimum phase type. Let the controlled system transfer function model be

$$P_n(s) = \frac{s - 1000}{s^3 + 7.2s^2 + 13.4s + 2.4}.$$

Following the same procedure as in Example 1, we again obtain the controller transfer function to be of the form

$$F_n(sI - A_n + B_nF_n + L_nC_n)^{-1}L_n,$$

where the state feedback gain and filter gain are as follows

$$F_n = [45.8000, 82.5780, 14.7394],$$
$$L_n = 10^4[-8.9951, -0.0934, -0.0003]',$$

and (A_n, B_n, C_n) is from a controller form realization of $P_n(s)$.

The Bode plot of loop transfer function with state feedback and with output feedback is shown in Fig. 5, respectively. The solid line is the Bode plot of target loop transfer function with state feedback. The Nyquist plot of the system with state feedback and observer based output feedback, respectively is shown in Fig. 6. The solid line is the Nyquist plot in the state feedback case and dashed line is the Nyquist plot when a suitable observer is used. It is shown that when the controlled process is of non-minimum phase, the

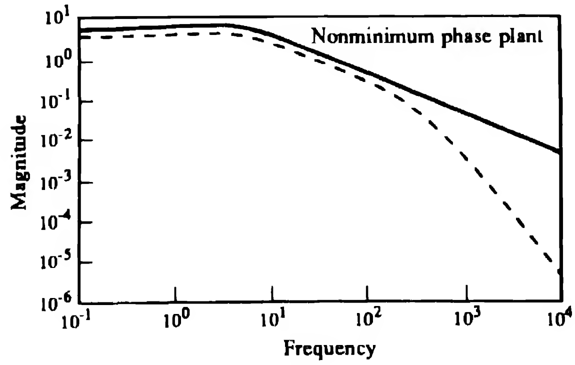


FIG. 5. Bode plot of target loop and recovery loop.

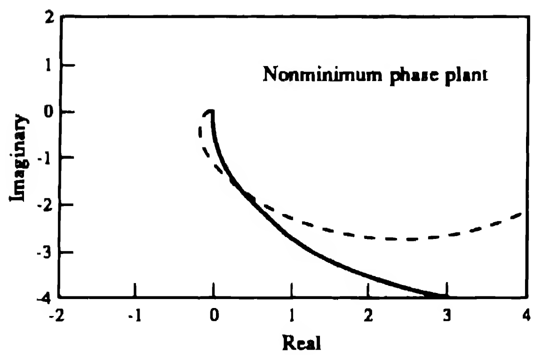


FIG. 6. Nyquist plot of systems with state and output feedback.

robustness property that the state feedback system has cannot be recovered by observer based control strategy, which also implies that LTR can not be achieved as good as desired. However, one can minimize the recovery error by applying the method we suggested here. Since we have taken care of the sensitivity when we design the system, from the Nyquist plot in Fig. 6, one can see that the output feedback system still has reasonably good stability robustness even though the system under control is of non-minimal phase.

5. *Conclusion*

It has been shown that the infinity norm of the sensitivity function of a closed loop system indicates stability robustness of the system in the sense of gain margin and phase margin. This result applies to both SISO systems and MIMO systems and therefore bridges the gap between characterizing stability robustness of SISO systems and of MIMO systems. Based on this result, we have suggested a way of designing an observer based control system which has 'good' stability robustness properties. If the controlled system is of minimal phase, we conclude that an observer based controller can be found with our approach such that the Nyquist plot of the output feedback system will only just go into the unit circle centered at $(-1, 0)$ in SISO case. If the controlled system is of non-minimal phase then one can never design an observer based control system to achieve this. We also suggest another way of doing an LQG/LTR design. It should be pointed out that the recovery quality is closely related to the stability robustness.

Acknowledgements—The first author wishes to thank Dr G. Gu for his very useful suggestions and comments. The authors also acknowledge the reviewers' valuable comments.

References

Anderson, B. D. O. and J. B. Moore (1971). *Linear Optimal Control.* Prentice-Hall, Englewood Cliffs, NJ.

Athans, M. (1986). A tutorial on the LQG/LTR method *Proc. of ACC*, pp. 1289–1296.

Boyd, S., V. Balakrishnan and P. Kabamba (1989). A bisection method for computing H^∞ norm of a transfer matrix and related problems. *Math. Control Signals Systems*, **2**, 207–219.

Doyle, J. C. (1978). Guaranteed margins for LQG regulators. *IEEE Trans. Aut. Control*, **AC-23**, 756–757.

Doyle, J. C. (1981). Multivariable feedback design: concepts for a classical/modern synthesis. *IEEE Trans. Aut. Control.* **AC-26**, 4–16.

Georgiou, T. T. and M. C. Smith (1990). Optimal robustness in the gap metric. *IEEE Trans. Aut. Control*, **35**, 673–686.

Lehtomaki, N. A., N. R. Sandell and M. Athans (1981). Robustness results in linear-quadratic Gaussian based multivariable control designs. *IEEE Trans. Aut. Control*, **AC-26**, 75–93.

Mageirou, E. F. and Y. C. Ho (1977). Decentralized stabilization via game theoretic methods. *Automatica*, **13**, 393–399.

Petersen, I. R. and C. V. Hollot (1988). High gain observers applied to problems in the stabilization of uncertain linear systems, disturbance attenuation and H^∞ optimization. *Int. J. of Adaptive Control and Signal Processing*, **2**, 347–369.

Vidyasagar, M and H. Kimura (1986). Robust controllers for uncertain linear multivariable systems. *Automatica*, **22**, 85–94.

Zhou, Kemin (1988). Robust control of uncertain systems and H_∞ optimization. Ph.D. thesis, Center for Control Science and Dynamic Systems, University of Minnesota

Automatica, Vol. 29, No. 2, pp. 485–489, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Stability Robustness of the Continuous-time LQG System Under Plant Perturbation and Noise Uncertainty*

J. S. LUO† and A. JOHNSON†

Key Words—Robustness; stability; linear optimal regulator; optimal control; state-space method.

Abstract—In this paper the robustness of the continuous-time LQG problem is studied. The system to be controlled is described by the usual state space formulation that includes the structure of plant parameter perturbations and noise uncertainty. A simple method is given to prove the saddle point inequality of the closed-loop system performance to cope with noise uncertainty. One sufficient condition, which is applicable to general linear systems, for system stability robustness under plant perturbation, is presented.

1. *Notation*

$A[\geq]B$ '$\geq$' is applied element by element to two matrices
$A \geq B$: $A - B \geq 0$ is a positive semidefinite matrix
$\|a\|$ is the norm of a vector a ($\|a\| = (a^Ta)^{1/2}$).
$\|A\|$ is the spectral norm of matrix A ($\|A\| = [\lambda_{\max}(AA^T)]^{1/2}$).
A^+ is a matrix obtained by replacing the entries of A with their absolute values.

2. *Introduction*

STANDARD LQG SYSTEM design is based on the assumption that the plant parameters and noise covariances are exactly known. But unfortunately both the plant parameters and the noise covariances are often approximations and subject to changes or are slowly varying. The existence of plant perturbation and noise uncertainty may degrade the system performance or even destabilize the controlled system.

Recently Chen and Dong (1989) proposed a state space approach to robust LQG control design for a continuous-time plant with norm bounded uncertainty in the parameters. Their first step, based on minimax theory developed in the works of Looze *et al.* (1983), Martin and Mintz (1983) and Verdu and Poor (1984), employed the minimax control scheme which minimizes the worst-case performance to cope with the noise uncertainty. In their second step, based on the Bellman–Gronwall inequality, they derived a sufficient condition for robust stability of the LQG continuous-time system under plant perturbations. In this paper we will follow the same procedure as adopted by Chen and Dong to solve the complete problem. After describing the conventional continuous-time LQG control problem in Section 3, we characterize the plant perturbation and noise uncertainty description in Section 4. In Section 5 we assume that the plant model is correct except that the noise covariances are uncertain. After the problem of noise uncertainty has been tackled, in Section 6 we derive a sufficient condition for the stability robustness of the continuous-time LQG problem under plant perturbation.

As opposed to Chen and Dong (1989), in this paper the plant perturbation and noise uncertainty are given in a more convenient form. In Section 5, instead of employing the minimax theorem we give a simple proof of the saddle point inequality of the closed-loop system performance, and in Section 6 we derive a sufficient condition for system stability robustness which can be applied to a general continuous-time linear system and which is less conservative than the conditions proposed by Chen and Dong (1989) and by Sobel *et al.* (1989). In Section 7 we give an example to illustrate that it is less conservative and this is followed with some conclusions in Section 8.

3. *Description of the conventional continuous-time LQG control problem*

Consider the continuous-time stochastic dynamic system described by:

$$\dot{x}(t) = Ax(t) + Bu(t) + w(t), \tag{3.1}$$

$$y(t) = Cx(t) + v(t). \tag{3.2}$$

It is assumed that the noise processes $w(t)$ and $v(t)$ are stationary white Gaussian with the following properties:

$$E\{w(t)\} = E\{v(t)\} = 0, \tag{3.3}$$

$$\mathrm{Cov}\,(w(t), w(s)) = W\delta(t-s), \quad W > 0, \tag{3.4}$$

$$\mathrm{Cov}\,(v(t), v(s)) = V\delta(t-s), \quad V > 0, \tag{3.5}$$

and $\{w(t)\}$ and $\{v(t)\}$ are independent of each other.

The performance index to be minimized is chosen as.

$$J = \lim_{t\to\infty} E\{x^T(t)Qx(t) + u^T(t)Ru(t)\}, \tag{3.6}$$

where

$$Q = Q^T \geq 0, \quad R = R^T > 0. \tag{3.7}$$

Throughout this paper the following assumptions are taken to hold (Johnson, 1985).

(1) $V > 0$.
(2) (A, C) is detectable.
(3) $(A, W^{1/2})$ is stabilizable.
(4) $R > 0$.
(5) (A, B) is stabilizable.
(6) $(A, Q^{1/2})$ is detectable.

Then a unique steady-state LQG regulator exists which produces a stable closed-loop and minimizes J in (3.6). The standard form of the steady-state LQG regulator is well known as follows

$$u(t) = -F\hat{x}(t), \tag{3.8}$$

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + H[y(t) - C\hat{x}(t)], \tag{3.9}$$

where

$$F = R^{-1}B^TP, \tag{3.10}$$

* Received 4 June 1991; revised 29 January 1992; revised 17 June 1992; received in final form 8 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor E. Kreindler under the direction of Editor H. Kwakernaak.

† Kramers Laboratory, Delft University of Technology, Prins Berhardlaan 6, 2628 BW Delft, The Netherlands.

$$0 = PA + A^TP + Q - PBR^{-1}B^TP, \tag{3.11}$$

$$H = \Sigma C^TV^{-1}, \tag{3.12}$$

$$0 = A\Sigma + \Sigma A^T + W - \Sigma C^TV^{-1}C\Sigma, \tag{3.13}$$

4. *The plant perturbation and noise uncertainty description*

Assume that the actual continuous-time LQG system could be described by:

$$\dot{x}(t) = [A + \Delta A(t)]x(t) + [B + \Delta B(t)]u(t) + w(t), \tag{4.1}$$

$$y(t) = [C + \Delta C(t)]x(t) + v(t), \tag{4.2}$$

where A, B and C are nominal parameters for LQG design. We suppose that bounds are available on the absolute values of the maximum variation in the elements of $\Delta A(t)$, $\Delta B(t)$ and $\Delta C(t)$ and that ΔA^m, ΔB^m and ΔC^m are defined as matrices with entries $(\Delta A^+_{ij})_{max}$, $(\Delta B^+_{ij})_{max}$ and $(\Delta C^+_{ij})_{max}$. Then the plant perturbations are represented as:

$$\{\Delta A(t): \Delta A^+(t)[\le]\, \Delta A^m\ \forall t \ge 0\}, \tag{4.3}$$

$$\{\Delta B(t): \Delta B^+(t)[\le]\, \Delta B^m\ \forall t \ge 0\}, \tag{4.4}$$

$$\{\Delta C(t): \Delta C^+(t)[\le]\, \Delta C^m\ \forall t \ge 0\}. \tag{4.5}$$

Meanwhile $w(t)$ and $v(t)$ in equations (4.1) and (4.2) are stationary and independent Gaussian white noise, defined as:

$$E\{w(t)\} = E\{v(t)\} = 0, \tag{4.6}$$

$$\begin{gathered}\mathrm{Cov}\,(w(t), w(s)) = (W + \Delta W)\,\delta(t-s),\\ W + \Delta W = (W + \Delta W)^T \ge 0, \quad W = W^T \ge 0,\end{gathered} \tag{4.7}$$

$$\begin{gathered}\mathrm{Cov}\,(v(t), v(s)) = (V + \Delta V)\,\delta(t-s),\\ V + \Delta V = (V + \Delta V)^T \ge 0, \quad V = V^T > 0,\end{gathered} \tag{4.8}$$

where W and V are nominal noise covariances for LQG design. Suppose that upper bounds are available for the actual covariances, i.e.

$$\{\Delta W: W + \Delta W \le W^m\}, \tag{4.9}$$

$$\{\Delta V: V + \Delta V \le V^m\}, \tag{4.10}$$

where $W^m \ge 0$ and $V^m > 0$ are symmetric matrices.

Figure 1 shows the actual continuous-time LQG system with the assumed plant perturbations and noise uncertainties

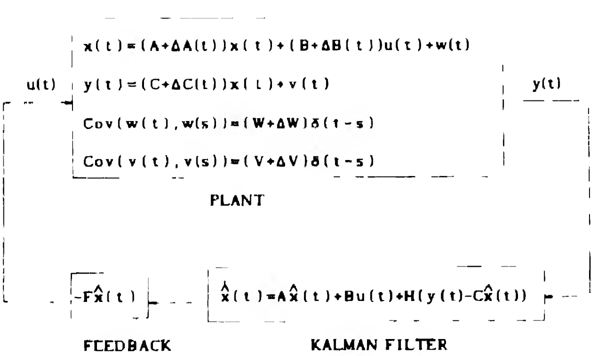


FIG. 1. The configuration of the closed-loop system.

5. *Robust continuous-time LQG design under noise uncertainty*

In this section we assume that the system parameter matrices A, B and C are correct (i.e. actual and design values are the same) and the noise uncertainty is defined as in (4.6)–(4.10). Under such assumptions, the optimal controller design assuming complete state information is still the standard optimal controller design described by (3.8), (3.10) and (3.11). However the Kalman Filter (KF) designed by using the nominal noise covariances may act very differently from what we would expect as a result of the noise covariances mismatch.

The standard KF design with nominal noise covariances is accomplished using formulas (3.9), (3.12) and (3.13), while the actual plant is given by (3.1) and (3.2) with $\{w(t)\}$ and $\{v(t)\}$ described by (4.6)–(4.10).

We define:

$$\tilde{x}^0(t) = x(t) - \hat{x}(t), \tag{5.1}$$

and we subtract (3.9) from (3.1) and put (5.1) into the result, so that we have

$$\dot{\tilde{x}}^0(t) = (A - HC)\tilde{x}^0(t) + w(t) - Hv(t). \tag{5.2}$$

Consequently the actual differential equation for the error covariance is:

$$\begin{aligned}\dot{\Sigma}^0(t) = {}&(A - HC)\Sigma^0(t) + \Sigma^0(t)(A - HC)^T\\ &+ W + \Delta W + H(V + \Delta V)H^T,\end{aligned} \tag{5.3}$$

where H is calculated from (3.12) and (3.13) using nominal noise covariances W and V. Sangsuk-Iam and Bullock (1988) proved that under assumptions (1)–(3) in Section 3, as $t \to \infty$ equation (5.3) always converges to some constant Σ^0 which satisfies the following Algebraic Lyapunov Equation (ALE):

$$\begin{aligned}0 = {}&(A - HC)\Sigma^0 + \Sigma^0(A - HC)^T\\ &+ W + \Delta W + H(V + \Delta V)H^T.\end{aligned} \tag{5.4}$$

Moreover, from (5.3) we notice that the estimation covariance is a monotonic increasing function of the actual noise covariance, i.e. if

$$W + \Delta W_1 \ge W + \Delta W_2 \quad \text{and} \quad V + \Delta V_1 \ge V + \Delta V_2,$$

and assuming

$$\Sigma^1(0) \ge \Sigma^2(0),$$

then

$$\Sigma^1(t) \ge \Sigma^2(t), \quad \forall t \ge 0$$

Now we consider the worst-case noise covariance

$$\dot{\Sigma}^m(t) = (A - HC)\Sigma^m(t) + \Sigma^m(t)(A - HC)^T + W^m + HV^mH^T$$

It converges to Σ^m which is the solution of the following ALE:

$$0 = (A - HC)\Sigma^m + \Sigma^m(A - HC)^T + W^m + HV^mH^T, \tag{5.5}$$

and this is the upper bound of the estimation covariance of the KF under any determined H. Using the notation $\Sigma(H, W, V)$ to emphasize that Σ is a function of H, W and V, we get

$$\Sigma(H, W + \Delta W, V + \Delta V) \le \Sigma(H, W^m, V^m). \tag{5.6}$$

It is a reasonable choice for us to take the worst-case noise covariances W^m and V^m to design the KF which minimize the upper bound of the KF performance, i.e.

$$\Sigma(H^mW^mV^m) \le \Sigma(HW^mV^m), \tag{5.7}$$

where H^m is the optimal gain matrix of the KF designed for the fixed pair of worst-case noise covariance W^m and V^m while H could be any other gain matrix. Combining (5.6) and (5.7), we have the following saddle point inequality.

$$\begin{aligned}\Sigma(H^m, W + \Delta W, V + \Delta V) &\le \Sigma(H^mW^mV^m)\\ &\le \Sigma(HW^mV^m).\end{aligned} \tag{5.8}$$

Now we investigate the LQG regulator, whose performance is measured by (3.6).

Theorem 5.1. The selection of W^m and V^m to minimize the upper bound of the KF performance as a strategy for robust KF design is also a strategy for robust LQG regulator design which minimizes the upper bound of the LQG regulator performance measured by (3.6).

Proof. See Appendix

Remark 5.1. The saddle point inequality for the maximal elements of the covariance uncertainties has been mentioned by Poor and Looze (1981) and Looze *et al.* (1983) for both the KF and LQG systems, respectively, as special cases of minimax theorems. Our proof here, when W^m and V^m are available, is simpler and clearer. Moreover, we give the

proof of the following property. If

$$\Omega_0 \geq \Omega_1$$

then

$$\operatorname{tr}\left(S \int_0^\infty e^{\bar{A}t}\Omega_0 e^{\bar{A}^T t}\, dt\right) \geq \operatorname{tr}\left(S \int_0^\infty e^{\bar{A}t}\Omega_1 e^{\bar{A}^T t}\, dt\right),$$

where Ω_0, Ω_1 and S are all symmetric positive semidefinite matrices. This property was used by Poor and Looze (1981) and Looze *et al.* (1983). In the authors' opinion, the proof of this property is important in understanding Theorem 5.1.

6. *A condition for stability robustness*

From the above discussion we may conclude that, under the stated assumptions, noise uncertainty has some influence on system performance but does not affect the asymptotic closed-loop system stability. However, the plant perturbations described by (4.1)–(4.5) may well cause instability of the system. In the following discussion, we derive a condition for stability robustness.

Putting (3.8) into (4.1) we obtain

$$\dot{x}(t) = [A \quad -BF]\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + [\Delta A(t) \quad -\Delta B(t)F]\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + w(t), \qquad (6.1)$$

where F is obtained from (3.10) and (3.11). Putting (3.8) and (4.2) into (3.9) produces:

$$\dot{\hat{x}}(t) = [HC \quad A - BF - HC]\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + [H\Delta C(t) \quad 0]\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + Hv(t), \qquad (6.2)$$

and combining (6.1) and (6.2) gives

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\hat{x}}(t) \end{bmatrix} = \begin{bmatrix} A & -BF \\ HC & A - BF - HC \end{bmatrix}\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} \Delta A(t) & -\Delta B(t)F \\ H\Delta C(t) & 0 \end{bmatrix}\begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & H \end{bmatrix}\begin{bmatrix} w(t) \\ v(t) \end{bmatrix}. \qquad (6.3)$$

Next we define the following

$$\bar{x}(t) \triangleq \begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix}, \quad G \triangleq \begin{bmatrix} A & -BF \\ HC & A - BF - HC \end{bmatrix},$$

$$\Delta G(t) \triangleq \begin{bmatrix} \Delta A(t) & -\Delta B(t)F \\ H\Delta C(t) & 0 \end{bmatrix},$$

$$\Delta G^+(t) = \begin{bmatrix} \Delta A^+(t) & \Delta B^+(t)F^+ \\ H^+\Delta C^+(t) & 0 \end{bmatrix},$$

and

$$\Delta G^m \triangleq \begin{bmatrix} \Delta A^m & \Delta B^m F^+ \\ H^+\Delta C^m & 0 \end{bmatrix}$$

Obviously

$$\Delta G(t)[\leq]\,\Delta G^+(t)[\leq]\,\Delta G^m, \quad \forall t \geq 0. \qquad (6.4)$$

For stability it is necessary and sufficient to consider the following system

$$\dot{\bar{x}}(t) = (G + \Delta G)\bar{x}(t). \qquad (6.5)$$

Now use the similarity transform M to obtain

$$\bar{x}(t) = Mz(t) \qquad (6.6)$$

Suppose M is the modal matrix to transform G to the Jordan matrix, i.e.

$$M^{-1}GM = \Lambda, \qquad (6.7)$$

where Λ is a Jordan matrix. Since $\|M\| < \infty$, if $\|z(t)\| \to 0$, implies

$$\|\bar{x}(t)\| \leq \|M\|\, \|z(t)\| \to 0. \qquad (6.8)$$

So instead of $\bar{x}(t)$, we could consider the stability of $z(t)$.

Applying the similarity transformation (6.6) to (6.5), we get

$$\dot{z}(t) = \Lambda z(t) + M^{-1}\Delta G(t)Mz(t), \qquad (6.9)$$

with the solution

$$z(t) = e^{\Lambda t}z(0) + \int_0^t e^{\Lambda(t-\tau)}M^{-1}\Delta G(\tau)Mz(\tau)\, d\tau. \qquad (6.10)$$

We know

$$(e^{\Lambda t})^+ [\leq] e^{\Lambda^0 t}, \quad \forall t \geq 0,$$

where

$$\Lambda^0 = \operatorname{Re}(\Lambda).$$

Taking absolute values on both sides of (6.10), we obtain

$$z^+(t) = \left[e^{\Lambda t}z(0) + \int_0^t e^{\Lambda(t-\tau)}M^{-1}\Delta G(\tau)Mz(\tau)\, d\tau\right]^+,$$

$$[\leq](e^{\Lambda t})^+z^+(0) + \int_0^t (e^{\Lambda(t-\tau)})^+(M^{-1})^+\,\Delta G^+(\tau)M^+z^+(\tau)\, d\tau,$$

$$[\leq]\, y_1(t) \triangleq (e^{\Lambda t})^+z^+(0) + \int_0^t (e^{\Lambda(t-\tau)})^+(M^{-1})^+\,\Delta G^+(\tau) \times M^+y_1(\tau)\, d\tau,$$

$$[\leq]\, y_2(t) \triangleq e^{\Lambda^0 t}z^+(0) + \int_0^t e^{\Lambda^0(t-\tau)}(M^{-1})^+\,\Delta G^m M^+ y_2(\tau)\, d\tau,$$

$$= e^{[\Lambda^0 + (M^{-1})^+\Delta G^m M^+]t}z^+(0). \qquad (6.11)$$

Obviously, if

$$\operatorname{Max} \operatorname{Re} \lambda_i[\Lambda^0 + (M^{-1})^+\,\Delta G^m M^+] < 0, \qquad (6.12)$$

we have

$$z^+(t) \to 0, \quad \forall z(0) < \infty \quad \text{as} \quad t \to \infty, \qquad (6.13)$$

i.e. the system (6.5) is asymptotically stable.

Because

$$\max \operatorname{Re} \lambda_i[\Lambda^0 + (M^{-1})^+\,\Delta G^m M^+] \leq \max \operatorname{Re} \lambda_i(\Lambda) + \operatorname{Max} \lambda_i[(M^{-1})^+\,\Delta G^m M^+], \qquad (6.14)$$

the sufficient stability robustness condition (6.12) which we have derived in this paper is less conservative than the condition of Sobel *et al.* (1986) which requires the right-hand side of inequality (6.14) to be negative. In the following Section 7 we give an example to illustrate that the condition (6.12) is less conservative than both the condition of Sobel *et al* (1989) and the condition of Chen and Dong (1989).

7. *Example*

We employ the same example as that used by Chen and Dong (1989) with a slight modification of ΔA for the convenience of comparison. The real continuous-time stochastic dynamic system is described by equations (4.1) and (4.2) with plant perturbations and noise uncertainty defined in Section 4, and

$$A = \begin{bmatrix} -1 & -3 \\ 1.5 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\Delta A = \begin{bmatrix} 1.1\cos(t) & 0 \\ 0 & 0 \end{bmatrix}, \quad \Delta B = \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix},$$

$$\Delta C(t) = \begin{bmatrix} 0.2\cos(t) & 0 \\ 0 & 0.1\sin(t) \end{bmatrix},$$

where ε is an uncertain value varying in $[-0.05 \quad 0.05]$

$$W = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} -0.1 & 0 \\ 0 & -0.7 \end{bmatrix} \leqq \Delta W \leq \begin{bmatrix} 0.1 & 0 \\ 0 & 0.7 \end{bmatrix},$$

and

$$\begin{bmatrix} -0.5 & 0 \\ 0 & -0.9 \end{bmatrix} \le \Delta V \le \begin{bmatrix} 0.5 & 0 \\ 0 & 0.9 \end{bmatrix}.$$

The performance index is defined as (3.6), and the weighting matrices are given as

$$Q = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad R = 1.$$

First we check the stability robust condition of Chen and Dong (1989)

$$\mu = \min\{\mu_1, \mu_2\} > m(2\alpha + 2\beta\,\|F\| + \gamma\,\|H\|), \qquad (7.1)$$

where

$$-\mu_1 \triangleq \text{Max Re}\,\lambda_i(A - BF) = -1.9408,$$

$$-\mu_2 \triangleq \text{Max Re}\,\lambda_i(A - HC) = -3.4879,$$

$$\|\Delta A(t)\| \le 1.1 = \alpha, \quad \|\Delta B(t)\| \le 0.05 = \beta,$$

$$\|\Delta C(t)\| \le 0.2 = \gamma,$$

$$F = [0.3345 \quad 0.5471] \quad \text{and} \quad H = \begin{bmatrix} 0.7414 & -0.0669 \\ -0.2191 & 0.2687 \end{bmatrix},$$

and $m = 1$ is a certain appropriate positive constant satisfying

$$\|e^{A^t}\| \le me^{-\mu t}, \quad \forall t \ge 0.$$

The result is

$$m(2\alpha + 2\beta\,\|F\| + \gamma\,\|H\|) = 2.4215.$$

Therefore the stability robust condition (7.1) of Chen and Dong (1989) is not satisfied.

Next we check the stability robustness condition of Sobel *et al.* (1989) and the stability robustness condition (6.12) we have derived in this paper. The feedback gain matrix F and KF gain matrix H are calculated from (3.10)–(3.13) using nominal parameters A, B, C and worst case noise covariances

$$W^m = \begin{bmatrix} 1.1 & 0 \\ 0 & 2.7 \end{bmatrix} \quad \text{and} \quad V^m = \begin{bmatrix} 2.5 & 0 \\ 0 & 1.9 \end{bmatrix}.$$

We get

$$F = [0.3345, 0.5471], \quad H = \begin{bmatrix} 0.7006 & -0.0735 \\ -0.2793 & 0.2578 \end{bmatrix}.$$

The modal matrix defined in (6.7) is

$$M = \begin{bmatrix} 0.1518 - 0.1171i & 0.1518 + 0.1171i & 0.6014 - 0.1209i & 0.6014 + 0.1209i \\ 0.0220 - 0.1031i & 0.0220 + 0.1031i & 0.0367 - 0.3497i & 0.0367 + 0.3497i \\ 0.6415 - 0.4015i & 0.6416 + 0.4015i & 0.6014 - 0.1209i & 0.6014 + 0.1209i \\ -0.4271 - 0.4437i & -0.4271 + 0.4437i & 0.0367 - 0.3497i & 0.0367 + 0.3497i \end{bmatrix}.$$

The definitions (4.3)–(4.5) lead to

$$\Delta A^m = \begin{bmatrix} 1.1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Delta B^m = \begin{bmatrix} 0.05 \\ 0 \end{bmatrix}, \quad \Delta C^m = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

To check the stability condition of Sobel *et al.* (1989)

$$-a + \pi[(M^{-1})^+ \Delta G^m M^+] < 0, \qquad (7.2)$$

where $a = -\text{Max Re}\,\lambda_i(\Lambda)$ and $\pi[(M^{-1})^+ \Delta G^m M^+]$ is the Perron–Frobenius radius of non-negative matrix $(M^{-1})^+ \Delta G^m M^+$. We obtain

$$-a + \pi[(M^{-1})^+ \Delta G^m M^+] = 0.2180.$$

The sufficient stability robustness condition (7.2) of Sobel *et al.* (1989) is not satisfied. But

$$\text{Max Re}\,\lambda_i[\Lambda^0 + (M^{-1})^+ \Delta G^m M^+] = -0.0140, \qquad (7.3)$$

the sufficient stability robustness condition (6.12) we have derived in this paper is satisfied, so the system is stable.

8. *Conclusion*

In this paper we have considered the continuous-time LQG problem. First we have given a direct and simple way to prove the saddle point inequality of the LQG system performance, to cope with the noise uncertainty, for robust LQG regulator design. Secondly a new sufficient condition of stability robustness for the continuous-time LQG regulator subject to plant perturbations has been presented in a convenient form of the uncertain plant parameters. The result is applicable to general linear systems. An example has shown that it is less conservative than two previous conditions presented by Chen and Dong (1989) and Sobel *et al.* (1989).

References

Chen, Bor-Sen and Dong Tay-Yuh (1989). LQG optimal control system design under plant perturbation and noise uncertainty: a state-space approach. *Automatica*, **25**, 431–436.

Johnson, A. (1985). *Process Dynamics Estimation and Control*. Peter Peregrinus, London.

Looze, D. P., H. V. Poor, K. S. Vastola and J. C. Darragh (1983). Minimax control of linear stochastic sysems with noise uncertainty. *IEEE Trans. Aut Control*, **AC-28**, 882–887.

Martin, C and M Mintz (1983). Robust filtering and prediction for linear system with uncertainty dynamics a game theoretic approach. *IEEE Trans Aut Control*, **AC-28**, 888–896.

Poor, V. and D P. Looze (1981). Minimax state estimation for linear stochastic systems with noise uncertainty *IEEE Trans. Aut. Control*, **AC-26**, 902–906

Sangsuk-Iam, S. and T. E Bullock (1988). Analysis of continuous-time Kalman filtering under incorrect noise covariances. *Automatica*, **24**, 659–669.

Sobel, K. M , S. S. Banda and H. H. Yeh (1989) Robust control for linear system with structured state space uncertainty *Int J. Control*, **50**, 1991–2004

Verdu, S. and H. V. Poor (1984). Minimax linear observers and regulators for stochastic system with uncertain second-order statistics. *IEEE Trans Aut Control*, **AC-28**, 499–511.

Appendix proof of Theorem 4.1

From (3.6)

$$\lim_{t\to\infty} E\{x^T(t)Qx(t) + u^T(t)Ru(t)\}$$

$$= \lim_{t\to\infty} E\{x^T(t)Qx(t) + (Fx(t) - F\bar{x}(t))^T R(Fx(t) - F\bar{x}(t))\}$$

$$= \lim_{t\to\infty} E\{x^T(t)(Q + F^TRF)x(t) - x^T(t)F^TRF\bar{x}(t)$$

$$- \bar{x}^T(t)F^TRFx(t) + \bar{x}^T(t)F^TRF\bar{x}(t)\}$$

$$= \lim_{t\to\infty} tr\begin{bmatrix} Q + F^TRF & -F^TRF \\ -F^TRF & F^TRF \end{bmatrix}$$

$$\times E\left\{\begin{bmatrix} x(t)x^T(t) & x(t)\bar{x}^T(t) \\ \bar{x}(t)x^T(t) & \bar{x}(t)\bar{x}^T(t) \end{bmatrix}\right\}. \qquad (A.1)$$

Then defining

$$\Xi(t) \triangleq E\left\{\begin{bmatrix} x(t) \\ \bar{x}(t) \end{bmatrix}\begin{bmatrix} x(t) \\ \bar{x}(t) \end{bmatrix}^T\right\} = E\left\{\begin{bmatrix} x(t)x^T(t) & x(t)\bar{x}^T(t) \\ \bar{x}(t)x^T(t) & \bar{x}(t)\bar{x}^T(t) \end{bmatrix}\right\}.$$

From (3.1), (3.8), (3.9), (5.1) and (5.2), we obtain

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\bar{x}}(t) \end{bmatrix} = \begin{bmatrix} A - BF & BF \\ 0 & A - HC \end{bmatrix}\begin{bmatrix} x(t) \\ \bar{x}(t) \end{bmatrix} + \begin{bmatrix} w(t) \\ w(t) - Hv(t) \end{bmatrix}. \qquad (A.2)$$

Next we define

$$\bar{A} \triangleq \begin{bmatrix} A - BF & BF \\ 0 & A - HC \end{bmatrix}, \quad S \triangleq \begin{bmatrix} Q + F^TRF & -F^TRF \\ -F^TRF & F^TRF \end{bmatrix},$$

and

$$\Omega \triangleq \begin{bmatrix} W+\Delta W & W+\Delta W \\ W+\Delta W & W+\Delta W+H(V+\Delta V)H^T \end{bmatrix}.$$

Under conditions (1)–(6) in Section 3, with F and H calculated from (3.10)–(3.13), all the eigenvalues of $\bar{A}$ lie strictly in the left half of complex plane, and therefore from (A.2) we have

$$\Xi(\infty) = \int_0^\infty e^{\bar{A}t}\Delta e^{\bar{A}^T t}\,dt \tag{A.3}$$

If we assume

$$\Delta W_1 - \Delta W_2 = \delta W = (\delta W)^T \geq 0,$$

and

$$\Delta V_1 - \Delta V_2 = \delta V = (\delta V)^T \geq 0,$$

then since

$$\begin{bmatrix} \delta W & \delta W \\ \delta W & \delta W \end{bmatrix} \geq 0 \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 0 & H\,\delta V H^T \end{bmatrix} \geq 0,$$

we obtain

$$\Delta\Omega \triangleq = \begin{bmatrix} \delta W & \delta W \\ \delta W & \delta W + H\,\delta V H^T \end{bmatrix} = \begin{bmatrix} \delta W & \delta W \\ \delta W & \delta W \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & H\,\delta V H^T \end{bmatrix} \geq 0, \tag{A.4}$$

and

$$\Delta\Omega = \Delta\Omega^T,$$

i.e. $\Delta\Omega$ is a positive semidefinite symmetric matrix. Similarly S is also a square symmetric positive semidefinite matrix. So there exists a square root of S satisfying

$$S^{1/2}(S^{1/2})^T = S^{1/2}S^{1/2} = S, \tag{A.5}$$

and therefore

$$\begin{aligned} \delta J \triangleq J_1 - J_2 &= \operatorname{tr} S\Xi_1(\infty) - \operatorname{tr} S\Xi_2(\infty) \\ &= \operatorname{tr} S(\Xi_1(\infty) - \Xi_2(\infty)) \\ &= \operatorname{tr} S\int_0^\infty e^{\bar{A}t}\Delta\Omega e^{\bar{A}^T t}\,dt \\ &= \operatorname{tr}\int_0^\infty S e^{\bar{A}t}\Delta\Omega e^{\bar{A}^T t}\,dt \\ &= \int_0^\infty \operatorname{tr} S e^{\bar{A}t}\Delta\Omega e^{\bar{A}^T t}\,dt \\ &= \int_0^\infty \operatorname{tr} S^{1/2} e^{\bar{A}t}\Delta\Omega e^{\bar{A}^T t} S^{1/2}\,dt \\ &= \int_0^\infty \operatorname{tr}\,(S^{1/2}e^{\bar{A}t})\,\Delta\Omega (S^{1/2}e^{\bar{A}t})^T\,dt. \end{aligned}$$

Because

$$(S^{1/2}e^{\bar{A}t})\,\Delta\Omega(S^{1/2}e^{\bar{A}t})^T \geq 0, \quad \forall t \geq 0,$$

we have

$$\operatorname{tr}\,(S^{1/2}e^{\bar{A}t})\,\Delta\Omega(S^{1/2}e^{\bar{A}t})^T \geq 0, \quad \forall t \geq 0,$$

and therefore

$$\int_0^\infty \operatorname{tr}\,(S^{1/2}e^{\bar{A}t})\,\Delta\Omega(S^{1/2}e^{\bar{A}t})^T\,dt \geq 0,$$

i.e.

$$J_1 \geq J_2, \tag{A.6}$$

which means that for any F and H, we will have the following inequality:

$$J(H, W+\Delta W, V+\Delta V) \leq J(H, W^m, V^m). \tag{A.7}$$

If we choose the worst-case noise covariances W^m and V^m to design the KF, the upper bound of the LQG regulator performance is minimized, i.e.

$$J(H^m, W^m, V^m) \leq J(H, W^m, V^m), \tag{A.8}$$

where H^m is computed from (3.12)–(3.13) using W^m and V^m and H could be any other matrix. Combining (A.7) and (A.8), we obtain the following saddle point inequality:

$$J(H^m, W+\Delta W, V+\Delta V) \leq J(H^m, W^m, V^m) \leq J(H, W^m, V^m). \tag{A.9}$$

Automatica, Vol. 29, No. 2, pp. 491–493, 1993
Printed in Great Britain

0005-1098/93 \$6.00 + 0.00

Brief Paper

Quadratic Stabilizability of Linear Uncertain Systems in Convex-bounded Domains*†

P. L. D. PERES,‡ J. C. GEROMEL‡ and J. BERNUSSOU§

Key Words—Control system design; robust control; linear systems; convex programming; uncertain systems.

Abstract—In this paper, a relationship is derived between quadratic stabilizability of linear systems with convex bounded uncertainty domains and the existence of a positive definite solution to a well defined set of Riccati equations. Both continuous and discrete-time models are investigated. For continuous-time systems, the results reported here are compared with the ones provided in the literature, where norm-bounded uncertainty is considered. A numerical example is included.

1. *Introduction*

It is well known that the Linear-Quadratic Problem plays a central role in linear control system design. In fact, among others, the controlled system exhibits several important properties as, for instance, the ones stated in Geromel and Cruz (1987) and Safonov and Athans (1987).

In this connection, the results provided in Petersen (1988) and Zhou and Khargonekar (1988) are of great importance. The authors showed that the existence of a positive definite solution to a modified Riccati equation is a necessary and sufficient condition for quadratic stabilization of uncertain systems by means of linear state feedback control. As a basic assumption, both papers deal with uncertain, continuous-time linear systems in norm-bounded uncertainty domains.

This paper provides similar results for linear uncertain systems with convex-bounded uncertainties. Both continuous and discrete-time systems are analysed and no additional structural assumptions are taken into account (e.g. matching conditions).

We claim (and we discuss this point by means of a simple example) that the convex-bounded uncertainty assumption can tackle most practical problems at least as well as the norm-bounded one. In fact, an important special case of convex-bounded uncertainty is the well-known interval matrices case.

2. *Preliminary considerations*

Let the uncertain linear system be defined as

$$\delta[x(t)] = Ax(t) + Bu(t), \tag{1}$$

where $x(\cdot) \in \Re^n$ is the state variable, $u(\cdot) \in \Re^m$ is the control variable and $\delta[\cdot]$ is a linear operator defined by $\delta[x(t)] \triangleq \dot{x}(t)$ for continuous-time systems and $\delta[x(t)] \triangleq x(t+1)$ for discrete-time ones. Associated with (1), we define the extended matrices (Barmish, 1983), $F \in \Re^{p\times p}$ and $G \in \Re^{p\times m}$, $p \triangleq n + m$:

$$F \triangleq \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}; \quad G \triangleq \begin{bmatrix} 0 \\ I \end{bmatrix}. \tag{2}$$

Matrices A and B, which define the model under consideration are not precisely known. In fact we assume that matrices A and B are uncertain and only their dimensions (n, m) are exactly known. The uncertainties are modeled in two different ways.

Norm-bounded uncertainty. For matrices D, E and Γ with appropriate dimensions, we define the uncertain domain $\mathcal{D}_N$ as being (Zhou and Khargonekar, 1988):

$$\mathcal{D}_N \triangleq \{F \in \Re^{p\times p} : F = F_0 + D\Gamma E;\ \|\Gamma\| \le 1\}. \tag{3}$$

Convex-bounded uncertainty. Given a set $\{F_i, i = 1 \cdots M\}$ of "extreme" matrices we define the uncertain domain $\mathcal{D}_C$ as being (Geromel *et al.*, 1991):

$$\mathcal{D}_C \triangleq \left\{F \in \Re^{p\times p} : F = \sum_{i=1}^{M} \lambda_i F_i;\ \lambda_i \ge 0;\ \sum_{i=1}^{M} \lambda_i = 1\right\}. \tag{4}$$

Relation (3) defines an ellipsoidal type of uncertainty domain and (4) a polyhedral one. The important case of interval matrices can be exactly modeled by $\mathcal{D}_C$ with appropriate choice of the extreme matrices $\{F_i, i = 1 \cdots M\}$. The same no longer holds for $\mathcal{D}_N$. Indeed, the matrix constraint $\|\Gamma\| \le 1$ introduces some dependencies among the parameters, leading in this case to an approximate representation of the uncertain domain.

In Zhou and Khargonekar (1988), it is proved that the continuous-time system (1) is quadratically stabilizable by linear control for all $F \in \mathcal{D}_N$ if and only if there exists $\epsilon > 0$ such that the modified Riccati equation

$$F_0'P + PF_0 - P(GR^{-1}G' - \epsilon DD')P + \frac{1}{\epsilon}E'E + Q = 0, \tag{5}$$

has a positive definite solution, where Q and R are arbitrary positive definite matrices. Assuming that a $P \in \Re^{p\times p}$ matrix satisfying (5) exists and partitioning its inverse in four blocks:

$$P^{-1} = W = \begin{bmatrix} W_1 & W_2 \\ W_2' & W_3 \end{bmatrix}, \tag{6}$$

where $W_1 \in \Re^{n\times n}$, $W_2 \in \Re^{n\times m}$ and $W_3 \in \Re^{m\times m}$, a linear state feedback stabilizing control as $u(t) = Kx(t)$ is readily determined from (Zhou and Khargonekar, 1988):

$$K = W_2'W_1^{-1} \tag{7}$$

In addition, with the Lyapunov function $v(x) = x'W_1^{-1}x$, the stability of the closed loop system can be verified for all $F \in \mathcal{D}_N$.

3. *Riccati equation for $\mathcal{D}_C$ uncertain domain*

In this section we present the main results of this paper. It is proved that, for continuous and discrete-time systems with

* Received 15 May 1991; revised 14 January 1992; revised 23 June 1992; received in final form 17 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor P. Dorato under the direction of Editor H. Kwakernaak.

† This research has been supported in part by grants from "Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP" and "Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq", Brazil.

‡ LAC-DT/FEE, UNICAMP, CP 6160, 13081-970, CAMPINAS/SP, Brazil.

§ LAAS/CNRS-7, avenue du Colonel Roche, 31077, Toulouse Cedex, France.

uncertainties belonging to $\mathscr{D}_C$, the concept of quadratic stabilizability via linear state feedback is equivalent to the existence of a positive definite solution to a set of algebraic Riccati equations. To this end, we first recall a result of Geromel *et al.* (1991), which states that the uncertain system (1) is quadratically stabilizable by linear control for all $F \in \mathscr{D}_C$ if and only if there exists $W = W' \in \Re^{p \times p}$ such that (see also Meilakhs (1978) for systems where only A is uncertain)

$$\begin{gathered} T'WT > 0, \\ T'\Theta_i(W)T < 0, \quad i = 1 \cdots M, \\ W \geqslant 0, \end{gathered} \tag{8}$$

where $T \in \Re^{p \times n}$ is an orthonormal matrix spanning the null space of G' and $\Theta_i(\cdot)$, $i = 1 \cdots M$ are linear matrix functions defined by $\Theta_i(W) \triangleq F_iW + WF_i'$ in the continuous-time case and $\Theta_i(W) \triangleq F_iWF_i' - W$ in the discrete-time one.

For $W \in \Re^{p \times p}$ partitioned as in (6), it is easy to verify that in the continuous-time case, the first $M + 1$ constraints in (8) do not depend explicitly on W_3. Consequently, without loss of generality, W_3 can be calculated in such way that $W > 0$. However, the same is not generally true for discrete-time systems.

Theorem 1. The continuous-time system (1) is quadratically stabilizable by linear control for all $F \in \mathscr{D}_C$ if and only if there exist positive definite matrices $Q_i \in \Re^{p \times p}$, $i = 1 \cdots M$ and $R \in \Re^{m \times m}$ such that the set of Riccati equations

$$F_i'P + PF_i - PGR^{-1}G'P + Q_i = 0, \quad i = 1 \cdots M, \tag{9}$$

have the same positive definite solution $P = P' > 0$.

Proof. Assuming that $P = P' > 0$ verifies (9) and defining $W = P^{-1}$, we have

$$\begin{aligned} T'\Theta_i(W)T &= T'(GR^{-1}G' - WQ_iW)T, \\ &< T'GR^{-1}G'T = 0. \end{aligned} \tag{10}$$

The last equality follows immediately from the definition of T. Since $W = P^{-1} > 0$, all conditions in (8) are fulfilled and the system is quadratically stabilizable $\forall F \in \mathscr{D}_C$.

Conversely, suppose that for some W, conditions (8) hold. From Finsler's Lemma (Geromel and Cruz, 1987) there exists $R = R' > 0$ such that

$$\Theta_i(W) < GR_i^{-1}G' < GR^{-1}G', \quad i = 1 \cdots M. \tag{11}$$

Defining positive definite matrices Q_i, $i = 1 \cdots M$ as

$$Q_i = W^{-1}[GR^{-1}G' - \Theta_i(W)]W^{-1}, \quad i = 1 \cdots M, \tag{12}$$

after simple algebraic calculations we can verify that $P = W^{-1}$ solves (9). This concludes the proof of the theorem proposed.

Some remarks are now in order. First, if P satisfies the previous theorem, then (7) provides a stabilizing gain for all $F \in \mathscr{D}_C$. Furthermore, to each model $F \in \mathscr{D}_C$ we can associate a Riccati equation, with the same solution $W = P^{-1}$. To see this, define

$$Q = \sum_{i=1}^{M} \lambda_i Q_i > 0, \tag{13}$$

and multiply (9) by $\lambda_i \in \Re_+$ and add terms to get

$$F'P + PF - PGR^{-1}G'P + Q = 0, \quad \forall F \in \mathscr{D}_C. \tag{14}$$

Note that this is possible since $\sum_{i=1}^{M} \lambda_i = 1$. Unfortunately, the result stated in Theorem 1 does not allow by itself the numerical determination of the P matrix. Indeed, it just brings to uncertain systems control design the properties of the Linear-Quadratic Problem (Safonov and Athans, 1987). For numerical purposes concerning the determination of $W = P^{-1}$, it appears to be better to solve directly (8) by means of methods described, for instance, in Boyd and Yang (1989) and Geromel *et al.* (1991).

We turn now our attention to the discrete-time case. In this sense, we need some results which can be easily proved. The first one is based on the convexity of the matrix function FWF' with respect to F for $W = W' \geq 0$. Consequently (see Geromel *et al.* (1991))

$$FWF' \leq \sum^{m} \lambda_i F_iWF_i', \quad \forall F \in \mathscr{D}_C. \tag{15}$$

The second one refers to the following equivalence for F, $V = V' > 0$ and $S = S' > 0$ being matrices with appropriate dimensions:

$$FVF' < S \Leftrightarrow F'S^{-1}F < V \tag{16}$$

Theorem 2. The discrete-time system (1) is quadratically stabilizable by linear control with $W > 0$, for all $F \in \mathscr{D}_C$ if and only if there exist positive definite matrices $Q_i \in \Re^{p \times p}$, $i = 1 \cdots M$ and $R \in \Re^{m \times m}$ such that the set of Riccati equations

$$F_i'PF_i - P - F_i'PG(R + G'PG)^{-1}G'PF_i + Q_i = 0, \quad i = 1 \cdots M, \tag{17}$$

have the same positive definite solution $P = P' > 0$.

Proof. Assuming that $P = P' > 0$ satisfies (17), using the matrix inversion lemma we get

$$\begin{aligned} P &> F_i'(P - PG(R + G'PG)^{-1}G'P)F_i \\ &= F_i'(P^{-1} + GR^{-1}G')^{-1}F_i, \quad i = 1 \cdots M \end{aligned} \tag{18}$$

On the other hand, defining $W = P^{-1}$ and using (16) in the above inequality, it follows that

$$T'\Theta_i(W)T < T'GR^{-1}G'T = 0. \tag{19}$$

Since $W = P^{-1}$ is positive definite, (8) holds and the discrete-time system (1) is quadratically stabilizable by the linear control law (7). Conversely, assume that (8) holds for some $W > 0$. From Finsler's Lemma, there exists a positive definite matrix $R \in \Re^{m \times m}$ such that

$$\Theta_i(W) < GR_i^{-1}G' < GR^{-1}G', \quad i = 1 \cdots M. \tag{20}$$

So, $F_iWF_i' < (W + GR^{-1}G')$, $i = 1 \cdots M$. Using again (16) it follows immediately that

$$W^{-1} > F_i'(W + GR^{-1}G')^{-1}F_i, \quad i = 1 \cdots M \tag{21}$$

Finally, defining the positive definite matrices Q_i as

$$Q_i \triangleq W^{-1} - F_i'(W + GR^{-1}G')^{-1}F_i, \quad i = 1 \cdots M, \tag{22}$$

and applying the matrix inversion lemma to (22) we verify that actually $P = W^{-1}$ solves the set of Riccati equations (17). This concludes the proof of the theorem proposed.

Once again, in the discrete-time case, a stabilizing state feedback gain is provided by (7). Furthermore, for any feasible model $F \in \mathscr{D}_C$ we have a Riccati equation associated to each of them. To see this, let us make use of equation (15) in (17) to get

$$\begin{aligned} \sum_{i=1}^{m} \lambda_i Q_i &= P - \sum_{i=1}^{M} \lambda_i F_i'(P^{-1} + GR^{-1}G')^{-1}F_i \\ &\leq P - F'(P^{-1} + GR^{-1}G')^{-1}F \\ &\triangleq Q. \end{aligned} \tag{23}$$

This means, that it is always possible to define for each $F \in \mathscr{D}_C$ a positive definite matrix Q satisfying (23) such that $P = P' > 0$ solves the Riccati equation:

$$\begin{aligned} Q &= P - F'(P^{-1} + GR^{-1}G')^{-1}F \\ &= P - F'PF + F'PG(R + G'PG)^{-1}G'PF, \quad \forall F \in \mathscr{D}_C \end{aligned} \tag{24}$$

For numerical purposes, the algorithm proposed in Geromel *et al.* (1991) can also handle the discrete-time case. If one exists, a $W \in \Re^{p \times p}$ satisfying (8) and solving also (24) can be readily determined.

For the sake of completeness, let us introduce a simple

example, consisting on the continuous-time model (1) where

$$A = \begin{bmatrix} 0 & \alpha - 1 \\ \beta & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \alpha \\ 1 - \beta \end{bmatrix}, \tag{25}$$

with α and β being the uncertain parameters. The goal is to determine the largest bound γ in such way $\forall \alpha = \beta \in \{|\alpha - 0.5| \leq \gamma\}$ the system remains quadratically stabilizable. The set $\mathcal{D}_N$ is defined by the matrices:

$$F_0 = \begin{bmatrix} 0 & -0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}, \quad D = \gamma \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$
$$E = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}; \quad \Gamma = \operatorname{diag}\left\{\frac{\alpha - 0.5}{\gamma}, \frac{\beta - 0.5}{\gamma}\right\}. \tag{26}$$

Using the numerical method proposed in Zhou and Khargonekar (1988) which iteratively tests the existence of a positive definite solution of (5), the maximum value for γ has been found: $\gamma_N = 0.271$. Associated to this value, (7) provides the stabilizing gain

$$K = [9.8683 \quad -50.0783].$$

At this point, it is important to add that for the given example the results of Zhou and Khargonekar (1988) are only sufficient for quadratic stabilizability. Indeed, the constraint $\|\Gamma\| \leq 1$ does not take into account the particular structure of matrix Γ which is in fact diagonal.

Now, using the convex programming method proposed in Geromel *et al.* (1991) we have solved (8) for two different situations:

$$\gamma = \gamma_N = 0.271, \quad K = [-0.0233 \quad -1.1010],$$
$$\gamma = \gamma_C = 0.363, \quad K = [-0.2884 \quad -5.9548]$$

The value of γ_C is the one which maximizes γ. As pointed out before, $\gamma_C > \gamma_N$ and from Theorem 1 we conclude that the Riccati equation (14) admits a positive definite solution for all $0 < \gamma \leq \gamma_C$. Clearly, this is an improvement when compared with the results of Zhou and Khargonekar (1988).

4. *Conclusions*

This paper establishes a relationship between the existence of a positive definite solution of a Riccati equation and the quadratic stabilizability of uncertain linear systems. It has been proved that, for both continuous and discrete-time uncertain models in convex-bounded domains, a Riccati equation exists whose solution is invariant over the space defined by the uncertain parameters.

This solution enables us to determine a stabilizing linear control law and brings to the uncertain system control design the well-known properties (e.g. robustness) of the Linear-Quadratic Problem solution.

References

Barmish, B. R. (1983). Stabilization of uncertain systems via linear control. *IEEE Trans. Aut. Control*, **AC-28**, 848–850.

Boyd, S. and Q. Yang (1989). Structured and simultaneously Lyapunov functions for system stability problems. *Int. J. Control*, **49**, 2215–2240.

Geromel, J. C. and J. J. da Cruz (1987). On the robustness of optimal regulators for nonlinear discrete-time systems. *IEEE Trans. Aut. Control*, **AC-32**, 703–710.

Geromel, J. C., P. L. D. Peres and J. Bernussou (1991). On a convex parameter space method for linear control design of uncertain systems. *SIAM J. Control and Optimization*, **29**, 381–402.

Meilakhs, A. M. (1978). Design of stable control systems subject to parametric perturbation. *Automatika i Telemekhanika*, **10**, 5–16.

Petersen, I. R. and C. V. Hollot (1986). A Riccati equation approach to the stabilization of uncertain linear systems. *Automatica*, **22**, 397–411.

Petersen, I. R. (1988). A stabilization algorithm for a class of uncertain linear systems. *System and Control Letters*, **8**, 351–357.

Safonov, M. G. and M. Athans (1987). Gain and phase margins for multiloop LQG regulators. *IEEE Trans. Aut. Control*, **AC-22**, 173–179.

Zhou, K. and P. P. Khargonekar (1988). Robust stabilization of linear systems with norm-bounded time-varying uncertainty. *Systems and Control Letters*, **10**, 17–20.

Automatica, Vol. 29, No. 2, pp. 495–498, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Transformation of Nonlinear Systems in Observer Canonical Form With Reduced Dependency on Derivatives of the Input*

T. PH. PROYCHEV† and R. L. MISHKOV†

Key Words—Nonlinear systems; nonlinear transformations; canonical forms; observers; observability; multivariable control systems.

Abstract—The transformation of nonlinear multi-input–multi-output systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$, $\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u})$ into an observer canonical form with reduced dependency on derivatives of the input is studied. Necessary and sufficient conditions for its existence and a straightforward algorithm for obtaining the canonical model are derived. The proposed method involves the solution of a nonlinear algebraic equation system and systems of first order linear partial differential equations. The nonlinear canonical form obtained permits global observer error linearization and it is a stage in the design of nonlinear observers. The method is illustrated by an example.

1. *Introduction*

THE SYNTHESIS OF nonlinear observers has received significant attention during the past few years. The basic results lie in the transformation of nonlinear systems into observer canonical forms admitting observer error linearization (Krener and Isidori, 1983; Bestle and Zeitz, 1983; Krener and Respondek, 1985; Isidori, 1989; Nijmeijer and van der Schaft, 1990). Transformation procedures for some classes of systems are developed (Zeitz, 1984, Keller, 1986a,b, 1987, Birk and Zeitz, 1988; Xia and Gao, 1989). The two-step transformation method proposed by Keller (1987) uses the generalized observer canonical form (GOCF) but it is not developed in the general case of nth order systems.

A nonlinear observer canonical form with reduced dependency on derivatives of the input and an approach for the transformation of nonlinear multi-input–multi-output (MIMO) systems are proposed here.

2. *Problem statement*

Consider the nonlinear MIMO system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{2.1a}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}), \tag{2.1b}$$

where $\mathbf{x}$ is a state n-vector, the input $\mathbf{u}$ is an r-vector, and the output $\mathbf{y}$ is an m-vector. The nonlinear vector functions $\mathbf{f}(\mathbf{x}, \mathbf{u})$ and $\mathbf{h}(\mathbf{x}, \mathbf{u})$ are assumed to be real and sufficiently differentiable in the considered domains of $\mathbf{x}$ and $\mathbf{u}$. The system can be time-variable if for example $u_i = t$. It is assumed to be locally observable with observability indices $n_i > 0$, $i = 1, 2, \ldots, m$, (Nijmeijer and van der Schaft, 1990) such that $n_1 + n_2 + \cdots + n_m = n$. The system

$$\begin{aligned}
y_1 &= h_1(\mathbf{x}, \mathbf{u}),\\
\dot{y}_1 &= Dh_1(\mathbf{x}, \mathbf{u}),\\
\ddot{y}_1 &= D^2h_1(\mathbf{x}, \mathbf{u}),\\
&\cdots\\
y_1^{(n_1-1)} &= D^{n_1-1}h_1(\mathbf{x}, \mathbf{u}),\\
y_2 &= h_2(\mathbf{x}, \mathbf{u}),\\
\dot{y}_2 &= Dh_2(\mathbf{x}, \mathbf{u}),\\
&\cdots\\
y_m^{(n_m-1)} &= D^{n_m-1}h_m(\mathbf{x}, \mathbf{u}),
\end{aligned} \tag{2.2}$$

(the observability map (Zeitz, 1984)) admits a solution of the type

$$\mathbf{x} = \mathbf{x}(\bar{\mathbf{y}}, \bar{\mathbf{u}}), \tag{2.3}$$

where $D = d/dt$ and

$$\bar{\mathbf{y}} = [\mathbf{y}^T, \dot{\mathbf{y}}^T, \ddot{\mathbf{y}}^T, \ldots, (\mathbf{y}^{(\nu-1)})^T]^T,$$
$$\bar{\mathbf{u}} = [\mathbf{u}^T, \dot{\mathbf{u}}^T, \ddot{\mathbf{u}}^T, \ldots, (\mathbf{u}^{(\nu)})^T]^T, \quad \nu = \max(n_1, n_2, \ldots, n_m).$$

The observer canonical form is

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z} + \mathbf{a}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}), \quad \mathbf{z}(0) = \mathbf{z}_0, \tag{2.4a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{z}, \tag{2.4b}$$

where

$$\mathbf{z} = [z_1, z_2, \ldots, z_{n_1}, z_{n_1+1}, z_{n_1+2}, \ldots, z_{n_1+n_2}, \ldots, z_n]^T,$$

$$\begin{aligned}\mathbf{a}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}) = [&a_1(\mathbf{y}, \mathbf{u}), a_2(\mathbf{y}, \mathbf{u}), \ldots, a_{n_1-1}\\
&\times (\mathbf{y}, \mathbf{u}), a_{n_1}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}), a_{n_1+1}(\mathbf{y}, \mathbf{u}),\\
&a_{n_1+2}(\mathbf{y}, \mathbf{u}), \ldots, a_{n_1+n_2}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}), \ldots, a_n(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})]^T,\end{aligned}$$

$\mathbf{A}$ and $\mathbf{C}$ are block-diagonal $n \times n$ and $m \times n$ matrices:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_m \end{bmatrix} \qquad \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} \\ \cdot & \cdot & & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_m \end{bmatrix}$$

$$\mathbf{A}_i = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}_{n_i} \qquad \mathbf{C}_i = [0 \quad \ldots \quad 0 \quad 1]_{1 \times n_i}$$

The proposed canonical form (2.4) differs from the GOCF for MIMO systems (Birk and Zeitz, 1988) in two ways—the output $\mathbf{y}$ depends linearly on the state $\mathbf{z}$ and the vector function $\mathbf{a}$ has a reduced dependency on the time derivatives of the input $\mathbf{u}$. This form permits an observer of the kind

$$\dot{\hat{\mathbf{z}}} = \mathbf{A}\hat{\mathbf{z}} + \mathbf{a}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}) + \mathbf{K}(\mathbf{y} - \hat{\mathbf{y}}),$$
$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{z}}.$$

* Received 3 May 1991; revised 3 January 1992, revised 11 May 1992; received in final form July 30 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor A. Isidori under the direction of Editor H. Kwakernaak.

† Control Systems Department, Technical University, Plovdiv, 18 Vasil Markov Str., 4000 Plovdiv, Bulgaria.

The observer error $\mathbf{e} = \mathbf{z} - \hat{\mathbf{z}}$ obeys the linear homogeneous equation $\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{KC})\mathbf{e}$ and the observer design can be achieved by pole assignment.

If the system (2.1) is transformable into the form (2.4), then the transformations

$$\mathbf{x} = \mathbf{T}(\mathbf{z}, \mathbf{u}) \quad \text{with} \quad \det(\partial \mathbf{T}/\partial \mathbf{z}) \neq 0, \tag{2.5a}$$

$$\mathbf{z} = \mathbf{T}^{-1}(\mathbf{x}, \mathbf{u}), \tag{2.5b}$$

exist and the equivalence condition

$$\mathbf{h}(\mathbf{x}, \mathbf{u}) = \mathbf{C}\mathbf{z}, \tag{2.6}$$

is satisfied for any $\mathbf{x}_0, \mathbf{u}, \mathbf{z}_0 = \mathbf{T}^{-1}(\mathbf{x}_0, \mathbf{u})$ or any $\mathbf{z}_0, \mathbf{u}$, and $\mathbf{x}_0 = \mathbf{T}(\mathbf{z}_0, \mathbf{u})$.

The transformation problem consists in finding the transformations (2.5) or the vector $\mathbf{a}(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})$ directly.

3. *Necessary and sufficient conditions for transformability*

Each of the m subsystems in the form (2.4) has a single ouput y_i, state vector $\mathbf{z}_i = [z_{j-n_i+1}, z_{j-n_i+2}, \ldots, z_j]^T$ and a nonlinear vector function $\mathbf{a}_i = [a_{j-n_i+1}, a_{j-n_i+2}, \ldots, a_j]^T$, in its model:

$$\dot{\mathbf{z}}_i = \mathbf{A}_i \mathbf{z}_i + \mathbf{a}_i(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}), \tag{3.1a}$$

$$y_i = z_j, \quad j = n_1 + n_2 + \cdots + n_i, \quad i = 1, 2, \ldots, m \tag{3.1b}$$

Let the system (2.1) be transformable into the form (2.4). The equivalence condition (2.6) holds and can be written componentwise:

$$h_i(\mathbf{x}, \mathbf{u}) = z_j, \quad j = \sum_{k=1}^{i} n_k, \quad i = 1, 2, \ldots, m. \tag{3.2}$$

The sequential differentiation of (3.2) with respect to time and substitution in accordance with (3.1) reads

$$D^{n_i} h_i(\mathbf{x}, \mathbf{u}) = D^{n_i-1} a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}) + D^{n_i-2} a_{j-1}(\mathbf{y}, \mathbf{u}) + \cdots + a_{j-n_i+1}(\mathbf{y}, \mathbf{u}). \tag{3.3}$$

The necessary conditions obtained are similar to the generalized characteristic equation used by Keller (1987), but the left hand sides are known through the original system (2.1) and the number of the unknown functions is one less.

The vector $\mathbf{x}$ can be substituted by the solution (2.3) and so the left hand side of (3.3) becomes a function of the vectors $\bar{\mathbf{y}}$ and $\bar{\mathbf{u}}$. Their components $\mathbf{y}, \dot{\mathbf{y}}, \ddot{\mathbf{y}}, \ldots, \mathbf{y}^{(n_i-1)}$ and $\mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}, \ldots, \mathbf{u}^{(n_i)}$ can be accepted as generalized arguments. The derivative $D^{n_i-1} a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})$ depends on all of them. It has such a structure that the arguments $\mathbf{y}^{(n_i-1)}$ and $\mathbf{u}^{(n_i)}$ are included linearly with coefficients $\partial a_j/\partial \mathbf{y}$ and $\partial a_j/\partial \dot{\mathbf{u}}$, respectively. The next derivative $D^{n_i-2} a_{j-1}(\mathbf{y}, \mathbf{u})$ does not depend on $\mathbf{y}^{(n_i-1)}, \mathbf{u}^{(n_i-1)}$ and $\mathbf{u}^{(n_i)}$. Thus the terms $D^{n_i-k} a_{j-k+1}(\mathbf{y}, \mathbf{u})$, $k = 2, 3, \ldots, n_i$ depend only on $\mathbf{y}, \dot{\mathbf{y}}, \ddot{\mathbf{y}}, \ldots, \mathbf{y}^{(n_i-k)}, \mathbf{u}, \dot{\mathbf{u}}, \ddot{\mathbf{u}}, \ldots, \mathbf{u}^{(n_i-k)}$ and also linearly on $\mathbf{y}^{(n_i-k)}$ and $\mathbf{u}^{(n_i-k)}$. Due to the structure of these derivatives and their specific dependence on the generalized arguments, the differentiating of (3.3) with respect to them yields the following equations as other necessary conditions.

$$\frac{\partial a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})}{\partial \mathbf{y}} = \frac{\partial h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}})}{\partial \mathbf{y}^{(n_i-1)}},$$

$$\frac{\partial a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})}{\partial \dot{\mathbf{u}}} = \frac{\partial h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}})}{\partial \mathbf{u}^{(n_i)}},$$

$$\frac{\partial a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})}{\partial \mathbf{u}} = \frac{\partial h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}})}{\partial \mathbf{u}^{(n_i-1)}} - (n_i - 1) \left(\dot{\mathbf{y}}^T \frac{\partial}{\partial \mathbf{y}} + \dot{\mathbf{u}}^T \frac{\partial}{\partial \mathbf{u}} + \ddot{\mathbf{u}}^T \frac{\partial}{\partial \dot{\mathbf{u}}} \right) \times \frac{\partial a_j}{\partial \dot{\mathbf{u}}}, \quad n_i > 1, \tag{3.4a}$$

$$\frac{\partial a_{j-n_i+k}(\mathbf{y}, \mathbf{u})}{\partial \mathbf{y}} = \frac{\partial}{\partial \mathbf{y}^{(k-1)}} \left\{ h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}}) - \sum_{s=2}^{n_i-k+1} D^{n_i-s+1} a_{j-s+2} \right\}, \tag{3.4b}$$

$$\frac{\partial a_{j-n_i+k}(\mathbf{y}, \mathbf{u})}{\partial \mathbf{u}} = \frac{\partial}{\partial \mathbf{u}^{(k-1)}} \left\{ h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}}) - \sum_{s=2}^{n_i-k+1} D^{n_i-s+1} a_{j-s+2} \right\}, \quad n_i > 2,$$

$$k = n_i - 1, n_i - 2, \ldots, 2, \quad j = \sum n_\nu, \quad i = 1, 2, \ldots, m,$$

where $\mathbf{y}^{(k)} = D^k \mathbf{y}$, $\mathbf{u}^{(k)} = D^k \mathbf{u}$ and $h_i^{(n_i)} = D^{n_i} h_i$.

If the system (2.1) is transformable into the special form (2.4) then the systems (3.4) are solvable and permit the corresponding unknown components of the vector function $\mathbf{a}$ to be computed. On composing the system (3.4a) the term $\partial a_j/\partial \dot{\mathbf{u}}$ in the third equation has to be substituted by the right hand side of the second equation.

It is well known that the systems of first order linear partial differential equations like (3.4) are integrable if the mixed partials commute. The system (3.4a) is integrable if and only if

$$\frac{\partial F_{j,s}(\mathbf{w})}{\partial w_k} = \frac{\partial F_{j,k}(\mathbf{w})}{\partial w_s}, \quad s, k = 1, 2, \ldots, m + 2r,$$

where $\mathbf{w} = [\mathbf{y}^T, \dot{\mathbf{u}}^T, \mathbf{u}^T]^T$, $F_{j,k}(\mathbf{w}) = \partial a_j/\partial w_k$ or

$$F_{j,1}(\mathbf{w}) = \frac{\partial h_i^{(n_i)}(.)}{\partial y_1^{(n_i-1)}}, \quad F_{j,2}(\mathbf{w}) = \frac{\partial h_i^{(n_i)}(.)}{\partial y_2^{(n_i-1)}}, \quad \ldots$$

$$F_{j,m+1}(\mathbf{w}) = \frac{\partial h_i^{(n_i)}(.)}{\partial u_1^{(n_i)}},$$

The integrability conditions for the systems (3.4b) are analogous.

Using the observability map (2.2), (2.3) and the necessary conditions (3.3) and (3.4), the following theorem can be formulated

Theorem 3.1. The system (2.1) is transformable in the canonical form (2.4) if and only if,

(i) there exist observability indices $n_i > 0$, $i = 1, 2, \ldots, m$, $n_1 + n_2 + \cdots + n_m = n$ for which the observability map (2.2) has the solution (2.3);

(ii) the systems of first order linear partial differential equations (3.4) are solvable with solutions of the kind $a_j(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}})$, $a_{j-n_i+k}(\mathbf{y}, \mathbf{u})$;

(iii) the functions

$$a_{j-n_i+1} = h_i^{(n_i)}(\bar{\mathbf{y}}, \bar{\mathbf{u}}) - \sum_{s=2}^{n_i} D^{n_i-s+1} a_{j-s+2}, \tag{3.5}$$

$$j = n_1 + n_2 + \cdots + n_i, \quad i = 1, 2, \ldots, m,$$

are non-null functions of the arguments $\mathbf{y}$ and $\mathbf{u}$.

Proof.

Necessity. The necessity of conditions (i) and (ii) follows immediately from the validity of the necessary conditions (3.4). The expressions (3.5) are a different record of the necessary conditions (3.3).

Sufficiency. Let conditions (i)–(iii) hold. From the conditions (ii) and (iii) it follows that the canonical model is completely defined. The sequential differentiating of (3.1b) and substituting in accordance with (3.1a) and (2.4b) yields the following expressions for the generalized variables $y_1, \dot{y}_1, \ddot{y}_1, \ldots, y_1^{(n_1-1)}, y_2, \dot{y}_2, \ddot{y}_2, \ldots, y_2^{(n_2-1)}, \ldots, y_m^{(n_m-1)}$:

$$y_i = z_j, \quad y_i^{(k)} = z_{j-k} + \sum_{s=1}^{k-1} D^{s-1} a_{j-k+s} \times (\mathbf{C}\mathbf{z}, \mathbf{u}) + D^{k-1} a_j(\mathbf{C}\mathbf{z}, \mathbf{u}, \dot{\mathbf{u}}), \tag{3.6}$$

$$i = 1, 2, \ldots, m, \; j = n_1 + n_2 + \cdots + n_i, \; k = 1, 2, \ldots, n_i - 1.$$

Obviously, the generalized variables are expressed as functions only of $\mathbf{z}$ and $\bar{\mathbf{u}}$. The substitution of expressions (3.6) in (2.3) immediately gives the straight transformation (2.5a). Therefore the system (2.1) is transformable in the form (2.4) and the equivalence condition (2.6) holds. If $\mathbf{Cz}$ is replaced by $\mathbf{h}(\mathbf{x}, \mathbf{u})$ in expressions (3.6) and the latter are

substituted in the left hand side of system (2.2) then the inverse transformation (2.5b) is directly determined in the form:

$$
\begin{aligned}
z_{j-k} &= D^k h_i(\mathbf{x}, \mathbf{u}) - \sum_{s=1}^{k-1} D^{s-1} a_{s+1}(\mathbf{h}(\mathbf{x}, \mathbf{u}), \mathbf{u}) \\
&\quad - D^{k-1} a_j(\mathbf{h}(\mathbf{x}, \mathbf{u}), \mathbf{u}, \dot{\mathbf{u}}) = T^{-1}_{j-k}(\mathbf{x}, \mathbf{u}), \qquad (3.7)\\
z_j &= h_i(\mathbf{x}, \mathbf{u}) = T^{-1}_j(\mathbf{x}, \mathbf{u}),
\end{aligned}
$$

$$i = 1, \ldots, m, \quad j = n_1 + n_2 + \cdots + n_i, \quad k = 1, 2, \ldots, n_i - 1.$$

The same form was derived by Keller (1987) for single output systems. If in (3.7) the vector $\mathbf{x}$ is represented by (2.3), then the vector $\mathbf{z}$ is expressed as a function of the generalized variables, too. So the transformations (2.5) exist and are invertible

4. *Transformation algorithm*

The necessary and sufficient conditions of Theorem 3.1 determine the following algorithm for transformation of system (2.1) into the form (2.4)

(i) Find an m-tuple of integers $n_i > 0$, $i = 1, 2, \ldots, m$, with $n_1 + n_2 + \cdots + n_m = n$, for which the Jacobian matrix of the observability map (2.2) with respect to the $\mathbf{x}$ vector (the observability matrix) has rank n. If such indices do not exist the system (2.1) is not observable and not transformable.

(ii) Find the solution (2.3) of the system (2.2) and replace it in the derivatives $D^{n_i} h_i(\mathbf{x}, \mathbf{u})$, $i = 1, 2, \ldots, m$, to turn them into functions of the generalized variables $\bar{\mathbf{y}}, \bar{\mathbf{u}}$.

(iii) For $i = 1, 2, \ldots, m$, if $n_i = 1$ then go to (iiib), else go to (iiia)

(iiia) Form the system (3.4a) and if it is solvable find the solution a_j, $j = \sum_{s=1}^{i} n_s$. If $n_i > 2$ then form the systems (3.4b) and if they are solvable find the solutions a_{j-n_i+k}, $k = n_i - 1, n_i - 2, \ldots, 2$. System (2.1) is nontransformable if any of the systems (3.4) is not solvable

(iiib) Determine the components a_{j-n_i+1} in accordance with (3.5). If they are non-null functions solely of the arguments $\mathbf{y}$ and $\mathbf{u}$ then the canonical model is determined, else system (2.1) is nontransformable into (2.4)

(iv) Form the expressions (3.6) and replace them in (2.3) to find the transformation $\mathbf{x} = \mathbf{T}(\mathbf{z}, \mathbf{u})$.

(v) Compute the inverse transformation $\mathbf{z} = \mathbf{T}^{-1}(\mathbf{x}, \mathbf{u})$ by the recurrent formulae following from (3.1):

$$
\begin{aligned}
z_j &= T^{-1}_j(\mathbf{x}, \mathbf{u}) = h_i(\mathbf{x}, \mathbf{u}), \\
z_{j-n_i+k} &= T^{-1}_{j-n_i+k}(\mathbf{x}, \mathbf{u}) - DT^{-1}_{j-n_i+k+1}(\mathbf{x}, \mathbf{u}) \\
&\quad - a_{j-n_i+k+1}(\mathbf{h}(\mathbf{x}, \mathbf{u}), \mathbf{u}, \dot{\mathbf{u}}),
\end{aligned}
$$

$$j = n_1 + n_2 + \cdots + n_i, \quad k = n_i - 1, n_i - 2, \ldots, 1,$$
$$i = 1, 2, \ldots, m$$

The integration of the systems (3.4) in step (iii) can be performed conveniently with zero initial conditions. Nonzero initial conditions can be used to scale the $\mathbf{z}$ vector. The algorithm requires the consecutive composition and solving of the first order systems of linear partial differential equations (3.4). As far as the solution found for the component a_k is used into the system of equations for a_{k-1}, the algorithm is recurrent.

5. *Example*

Consider the system

$$
\begin{aligned}
\dot{x}_1 &= 2\sqrt{x_1}(x_1 + x_4^2 + u_2) - 4x_1x_4, \\
\dot{x}_2 &= (x_3 + 1)(x_4 - \sqrt{x_1}) + (\sqrt{x_1} - x_4)u_1, \\
\dot{x}_3 &= x_2 + x_3 - u_1, \\
\dot{x}_4 &= (\sqrt{x_1} - x_4)^2 + (x_3 - u_1)^2 + \sqrt{x_1}, \\
y_1 &= x_4 - \sqrt{x_1}, \\
y_2 &= x_3 - u_1.
\end{aligned}
$$

On $n_1 = n_2 = 2$ the observability map (2.2) takes the form

$$
\begin{aligned}
y_1 &= x_4 - \sqrt{x_1}, \\
\dot{y}_1 &= (\sqrt{x_1} - x_4)^2 + (x_3 - u_1)^2 + \sqrt{x_1} \\
&\quad - x_1 - x_4^2 - u_2 + 2\sqrt{x_1}x_4, \\
y_2 &= x_3 - u_1, \\
\dot{y}_2 &= x_2 + x_3 - u_1 - \dot{u}_1.
\end{aligned}
$$

Its Jacobian matrix has rank 4 and the solution (2.3) being:

$$
\begin{aligned}
x_1 &= (\dot{y}_1 - y_2^2 + u_2)^2, \\
x_2 &= \dot{y}_2 - y_2 + \dot{u}_1, \\
x_3 &= y_2 + u_1, \\
x_4 &= \dot{y}_1 - y_2^2 + u_2 + y_1.
\end{aligned}
$$

The derivatives $h_i^{(n_i)}(\mathbf{x}, \mathbf{u})$ are

$$
\begin{aligned}
h_1^{(2)} &= x_1 + x_4^2 + u_2 - 2\sqrt{x_1}\,x_4 + 2(x_3 - u_1) \\
&\quad \times (x_2 + x_3 - u_1) - 2(x_3 - u_1)\dot{u}_1 - \dot{u}_2, \\
h_2^{(2)} &= (x_3 + 1)(x_4 - \sqrt{x_1}) + (\sqrt{x_1} - x_4) \\
&\quad \times u_1 + x_2 + x_3 - u_1 - \dot{u}_1 - \ddot{u}_1,
\end{aligned}
$$

and after substitution of the solution (2.3) take the form

$$
\begin{aligned}
h_1^{(2)} &= 2y_2\dot{y}_2 - \dot{u}_2 + y_1^2 + u_2, \\
h_2^{(2)} &= \dot{y}_2 - \ddot{u}_1 + y_1 + y_1y_2.
\end{aligned}
$$

The systems (3.4a) read

$$
\frac{\partial a_2}{\partial \mathbf{y}} = \begin{bmatrix} 0 \\ 2y_2 \end{bmatrix}, \quad \frac{\partial a_2}{\partial \dot{\mathbf{u}}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \frac{\partial a_2}{\partial \mathbf{u}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix},
$$

$$
\frac{\partial a_4}{\partial \mathbf{y}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \frac{\partial a_4}{\partial \dot{\mathbf{u}}} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \frac{\partial a_4}{\partial \mathbf{u}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
$$

These systems are solvable and have solutions $a_2 = y_2^2 - u_2$, $a_4 = y_2 - \dot{u}_1$. From (3.5) there follows

$$
\begin{aligned}
a_1 &= h_1^{(2)}(\bar{\mathbf{y}}, \bar{\mathbf{u}}) - Da_2(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}) = y_1^2 + u_2, \\
a_3 &= h_2^{(2)}(\bar{\mathbf{y}}, \bar{\mathbf{u}}) - Da_4(\mathbf{y}, \mathbf{u}, \dot{\mathbf{u}}) = y_1 + y_1y_2
\end{aligned}
$$

These functions depend solely on $\mathbf{y}$ and $\mathbf{u}$. Then the canonical model is

$$
\begin{aligned}
\dot{z}_1 &= z_2^2 + u_2, \\
\dot{z}_2 &= z_1 + z_4^2 - u_2, \\
\dot{z}_3 &= z_2 + z_2z_4, \\
\dot{z}_4 &= z_3 + z_4 - \dot{u}_1, \\
y_1 &= z_2, \\
y_2 &= z_4.
\end{aligned}
$$

The straight and inverse transformations are computed in accordance with steps (iv) and (v) of the algorithm:

$$
\begin{aligned}
\mathbf{x} &= \mathbf{T}(\mathbf{z}, \mathbf{u}) = [z_1^2, z_3, z_4 + u_1, z_1 + z_2]^T, \\
\mathbf{z} &= \mathbf{T}^{-1}(\mathbf{x}, \mathbf{u}) = [\sqrt{x_1}, x_4 - \sqrt{x_1}, x_2, x_3 - u_1]^T.
\end{aligned}
$$

6. *Conclusions*

The method proposed and the two-step transformation method of Keller (1987) are essentially similar. Due to the special kind of the introduced canonical form, the transformation method considered has been developed as a straightforward algorithm for nth order nonlinear MIMO systems. The class of the transformable systems is restricted by the reduced dependency on time derivatives of the input and the linear output equation of the observer canonical model.

References

Bestle, D. and M. Zeitz (1983). Canonical form observer design for non-linear time-variable systems. *Int. J. Control*, **38**, 419–431.

Birk, J. and M. Zeitz (1988). Extended Luenberger observer for non-linear multivariable systems. *Int. J. Control*, **47**, 1823–1836

Isidori, A. (1989). *Nonlinear Control Systems*. Springer, Berlin.

Keller, H. (1986a). *Automatisierungstechnik*, **34**, 271–274, 326–331.

Keller, H. (1986b). *Entwurf Nichtlinearer Beobachter Mittels Normalformen*. Verlag des Vereins Deutscher Ingenieure, Dusseldorf.

Keller, H. (1987). Non-linear observer design by transformation into a generalized observer canonical form. *Int. J. Control*, **46**, 1915–1930.

Krener, A. J. and A. Isidori (1983). Linearization by output injection and non-linear observers. *Syst. and Contr. Lett.*, **3**, 47–52.

Krener, A. J. and W. Respondek (1985). Non-linear observers with linearizable error dynamics. *SIAM J. Control and Opt.*, **23**, 197–216.

Nijmeijer, H. and A. J. van der Schaft (1990). *Nonlinear Dynamical Control Systems*. Springer, Berlin.

Xia, X. and W. Gao (1989). Non-linear observer design by observer error linearization. *SIAM J. Contr. and Opt.*, **27**, 199–216.

Zeitz, M. (1984). Observability canonical (phase-variable) form for non-linear time-variable systems. *Int. J. Syst. Sci.*, **15**, 949–958.

Automatica, Vol. 29, No 2, pp 499–505, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

A Nonlinear Fuzzy Controller with Linear Control Rules is the Sum of a Global Two-dimensional Multilevel Relay and a Local Nonlinear Proportional-integral Controller*

HAO YING†

Key Words—Control system analysis; fuzzy control; nonlinear control systems; PID control; relay control.

Abstract—The author analytically proves that a nonlinear fuzzy controller with linear control rules and N members for input fuzzy sets is the sum of a global two-dimensional multilevel relay and a local nonlinear proportional-integral (PI) controller which adjusts the control action generated by the global multilevel relay. As N increases, the resolution of the global multilevel relay is enhanced but the role of the local nonlinear PI controller in total control action is decreased. As N approaches ∞, the global multilevel relay approaches a regular linear PI controller while the control action from the local nonlinear PI controller approaches zero. The role of the global multilevel relay and the local nonlinear PI controller in total control action is quantitatively described, as is the degree of nonlinearity of the fuzzy controllers with different N.

1. *Introduction*

TO ADVANCE FUZZY control technique, sound theory needs to be developed. The author believes that one way to develop such theory is to analytically investigate structures of fuzzy controllers and relate the structures to nonfuzzy control theory. Such relations will provide solid frameworks for analytically solving many important but previously difficult problems in fuzzy control technique, such as stability, by utilizing abundant well-developed and powerful nonfuzzy control techniques.

To reveal structures of fuzzy controllers and link the structures with nonfuzzy control theory, a novel method was initially developed (Ying, 1987), presented (Ying *et al.*, 1988) and published (Ying *et al.*, 1990). The work showed that the simplest possible nonlinear fuzzy controller with two members for the input fuzzy sets, "error" and "rate change of error" ("rate" for short) was equivalent to a regular linear PI controller when a linear defuzzification algorithm was used or to a nonlinear PI controller when a nonlinear defuzzification algorithm was used. Using this method, the results on the linear properties of the fuzzy controller were generalized to fuzzy controllers with more members for the input fuzzy sets and different fuzzy logic, first by Siler and Ying (1989) and then by Buckley and Ying (1990) and Buckley (1989a). Moreover, the Limit Theorems for linear fuzzy control rules were developed (Buckley and Ying, 1989) and extended to multiple-input–multiple-output fuzzy controllers (Buckley, 1990). Following the generalization of the results on the linear properties, the results on the nonlinear aspects of the fuzzy controller were also mathematically generalized to the fuzzy controller with more members for the input fuzzy sets, first by Buckley (1989b) and then by Wang *et al.* (1990).

In this paper, a nonlinear fuzzy controller with linear control rules is first defined. The author then analytically derives the explicit structure of the fuzzy controller and relates the resultant structure to the multilevel relay and PI controller of nonfuzzy control theory.

2. *Theoretical analysis of structure of the nonlinear fuzzy controller*

2.1. *Components of the nonlinear fuzzy controller.* If T denotes sampling period and nT (n is a positive integer) denotes sampling time, then the scaled inputs at sampling time nT are

$$e^* = GE \cdot e(nT) = GE[y(nT) - \text{setpoint}], \qquad (2.1)$$

$$r^* = GR \cdot r(nT) = GR[e(nT) - e(nT - T)], \qquad (2.2)$$

where $e(nT)$, $r(nT)$ and $y(nT)$ designate crisp unscaled error, rate, and process output at sampling time nT, respectively, and $e(nT - T)$ specifies crisp unscaled error at sampling time $(n - 1)T$. GE and GR are the scalars for the crisp error and rate.

Let the number of members of the fuzzy sets "error" and "rate" be the same and the membership functions be identical. This condition can easily be met, since if the number of members differs, some members can be added to the smaller set to attain equality. Assume there are $J(J \geq 1)$ members for positive "error" ("rate"), J members for negative "error" ("rate") and one member for zero "error" ("rate"). Therefore, there are a total of

$$N = 2J + 1 \geq 3, \qquad (2.3)$$

members for the fuzzy set "error" ("rate"). Index systems

$$\{E_{-J}, E_{-J+1}, \ldots, E_{-1}, E_0, E_1, \ldots, E_{J-1}, E_J\},$$

and

$$\{R_{-J}, R_{-J+1}, \ldots, R_{-1}, R_0, R_1, \ldots, R_{J-1}, R_J\}, \qquad (2.4)$$

are adopted to establish relationships between the indexes and the names of the members of the fuzzy sets "error" and "rate". E_i represents a member of the fuzzy set "error" and R_i represents a member of the fuzzy set "rate". The positive indexes specify the members for positive error (rate), the negative indexes denote the members for negative error (rate) and the index 0 corresponds to the zero error (rate) of the fuzzy sets. The membership functions corresponding to

* Received 11 July 1991; revised 10 March 1992; received in final form 8 April 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor M. Ikeda under the direction of Editor A. P. Sage.

† Department of Physiology and Biophysics, Biomedical Engineering Center and Office of Academic Computing, University of Texas Medical Branch, Galveston, TX 77550, U.S.A.

the members in (2.4) are expressed as:

$$\{\mu_{-J}(x), \mu_{-J+1}(x), \ldots, \mu_{-1}(x), \mu_0(x), \mu_1(x), \ldots, \mu_{J-1}(x), \mu_J(x)\}. \quad (2.5)$$

Denote the central value of the membership function $\mu_i(x)$ as λ_i and define $\lambda_{-J} = -L$, $\lambda_0 = 0$, and $\lambda_J = L$. Also, let the space between the central values of two adjacent members be equal. Then the space S is:

$$S = \frac{L}{J}, \quad (2.6)$$

and consequently the central value of $\mu_i(x)$ is $\lambda_i = i \cdot S$. It is obvious that the base of each member is $2S$. It should be noted that the equality of the bases does not imply loss of generality because the bases of the members of "error" and "rate" are different with respect to the actual unscaled inputs, $e(nT)$ and $r(nT)$.

The membership function $\mu_i(x)$ in this study is the commonly-used triangular-shaped membership function satisfying the following conditions:

(1) For $i = -J+1, -J+2, \ldots, J-2, J-1$,

$$\mu_i(x) = \frac{1}{S}[x - (i-1)S], \quad \text{if } x \in [(i-1)S, iS],$$

$$\mu_i(x) = -\frac{1}{S}[x - (i+1)S], \quad \text{if } x \in [iS, (i+1)S],$$

$$\mu_i(x) = 0, \quad \text{if } x \notin [(i-1)S, (i+1)S]$$

(2) For $i = J$ or $i = -J$,

$$\mu_J(x) = \frac{1}{S}[x - (J-1)S], \quad \text{if } x \in [(J-1)S, JS],$$

$$\mu_J(x) = 1, \quad \text{if } x \in [JS, +\infty)$$
$$\text{and } \mu_J(x) = 0, \quad \text{if } x \notin [(J-1)S, +\infty),$$

$$\mu_{-J}(x) = -\frac{1}{S}[x - (-J+1)S], \quad \text{if } x \in [-JS, (-J+1)S],$$

$$\mu_{-J}(x) = 1, \quad \text{if } x \in (-\infty, JS]$$
$$\text{and } \mu_{-J}(x) = 0, \quad \text{if } x \notin (-\infty, (-J+1)S].$$

It is obvious that

$$\mu_i(x) + \mu_{i+1}(x) = 1, \quad x \in (-\infty, +\infty). \quad (2.7)$$

Figure 1 shows an example of such a membership function with $N = 7$ ($J = 3$) and $S = 5$. In this paper, $\mu_i(e^*)$ is denoted as the membership for E_i and $\mu_i(r^*)$ as the membership for R_i.

Assume there are $2N - 1$ (or $4J + 1$) members in the fuzzy set "output". Among these, $2J$ members are for positive "output," $2J$ members are for negative "output" and one member is for zero "output". Using the index system (2.5), the members of the fuzzy set "output" can be described by

$$\{U_{-2J}, U_{-2J+1}, \ldots, U_{-1}, U_0, U_1, \ldots, U_{2J-1}, U_{2J}\}. \quad (2.8)$$

The central values of the members of the fuzzy set "output" are designated as γ_i. Let $\gamma_{-2J} = -H$, $\gamma_0 = 0$ and $\gamma_{2J} = H$. Further, let the space V between the central values of two adjacent members be equal. Therefore the space is

$$V = \frac{H}{2J} = \frac{H}{N} \quad (2.9)$$

and the ith central value can be written as

$$\gamma_i = i \cdot V = \frac{i \cdot H}{N-1}. \quad (2.10)$$

For the fuzzy set "output," the author requires: (1) the membership function be symmetrical about its central value; and (2) the shape of the membership functions of all the members be the same.

It is necessary to use N^2 fuzzy control rules to cover all the possible combinations of N members of the fuzzy set "error" and N members of the fuzzy set "rate". In the study, the fuzzy control rules must comply with the following rule:

IF "error" is E_i and "rate" is R_j
THEN "output" is $U_{-(i+j)}$. (2.11)

In other words, the index of the member of "output" is always equal to the negative sum of the indexes of the members of "error" and the members of "rate". Such a control rule is called here a linear control rule. To illustrate this rule, take $N = 5$ as an example. If the members of the input fuzzy sets are {negative medium (NM), negative small (NS), zero (ZO), positive small (PS), positive medium (PM)} and are represented as $\{E_{-2}, E_{-1}, E_0, E_1, E_2\}$ and $\{R_{-2}, R_{-1}, R_0, R_1, R_2\}$, then the corresponding nine members of the fuzzy set "output" can be indexed as $\{U_{-4}, U_{-3}, U_{-2}, U_{-1}, U_0, U_1, U_2, U_3, U_4\}$ which may be interpreted as {negative very large (NVL), negative large

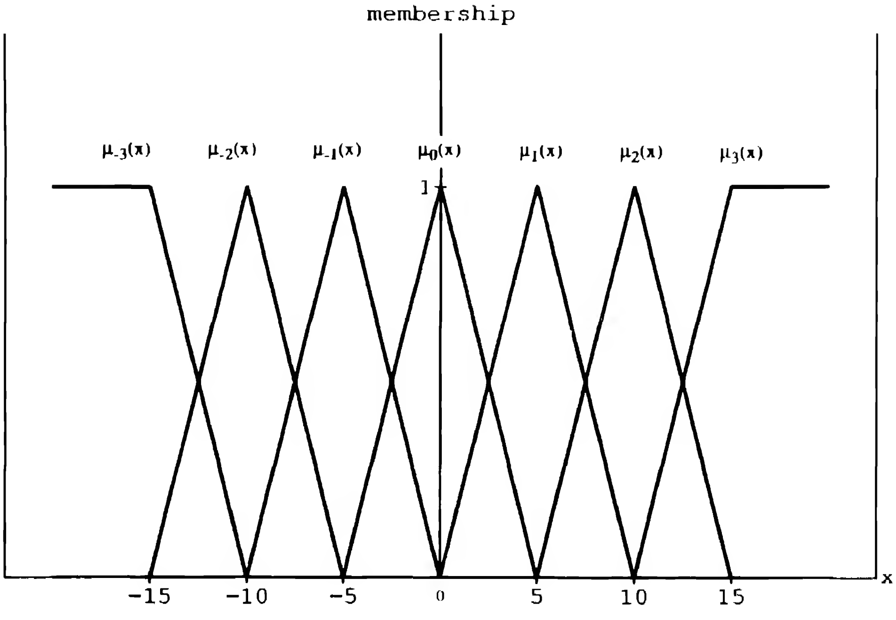


FIG. 1. An example of a triangular-shaped membership function with seven members ($N = 7$). The space between the central value of two adjacent members is 5 ($S = 5$).

TABLE 1. AN EXAMPLE TO SHOW HOW TO CONSTRUCT 25 FUZZY CONTROL RULES ACCORDING TO THE RULE (2.11) WHEN $N = 5$

	E_{-2}(NM)	E_{-1}(NS)	E_0(ZO)	E_1(PS)	E_2(PM)
R_{-2}(NM)	U_4(PVL)	U_3(PL)	U_2(PM)	U_1(PS)	U_0(ZO)
R_{-1}(NS)	U_3(PL)	U_2(PM)	U_1(PS)	U_0(ZO)	U_{-1}(NS)
R_0(ZO)	U_2(PM)	U_1(PS)	U_0(ZO)	U_{-1}(NS)	U_{-2}(NM)
R_1(PS)	U_1(PS)	U_0(ZO)	U_{-1}(NS)	U_{-2}(NM)	U_{-3}(NL)
R_2(PM)	U_0(ZO)	U_{-1}(NS)	U_{-2}(NM)	U_{-3}(NL)	U_{-4}(NVL)

(NL), NM, NS, ZO, PS, PM, positive large (PL), positive very large (PVL)}. The corresponding 25 fuzzy control rules satisfying the rule (2.11) are shown in Table 1.

Zadeh fuzzy logic AND is used to execute the IF side of the fuzzy control rule. That is

$$\mu(i, j) = \mathrm{Min}\,(\mu_i(e^*), \mu_j(r^*)), \qquad (2.12)$$

where $\mu(i, j)$ is the membership of the member of the fuzzy set "output" obtained when E_i and R_j are used in the IF side. Because the membership function of "output" is symmetrical about its central value, the central value of the member $U_{(i+j)}$, $\gamma_{(i+j)}$, and the resultant membership from the IF side, namely $\mu(i, j)$, are used to calculate the THEN side of the fuzzy control rule, i.e.

$$v(i, j) = \mu(i, j)\cdot\gamma_{(i+j)} = -\mathrm{Min}\,(\mu_i(e^*), \mu_j(r^*))\cdot(i+j)V, \qquad (2.13)$$

where $v(i, j)$ is the incremental control output contributed by the fuzzy control rule (2.11). If more than one membership results, say μ_1 and μ_2, from executing two different fuzzy control rules, Lukasiewicz fuzzy logic OR is used to get combined membership, μ, because the conditions being ORed are maximally negatively correlated. That is

$$\mu = \mathrm{Min}\,(\mu_1 + \mu_2, 1) \qquad (2.14)$$

Recall that the shapes of the membership functions of "output" were required to be the same. In the defuzzification process, therefore, the contribution from the members of "output" in the THEN side of the fuzzy control rules should be weighted by their memberships calculated from the IF side. Consequently, the scaled crisp incremental output $GU\cdot\Delta u(nT)$ can be calculated by the following defuzzification algorithm

$$GU\cdot\Delta u(nT) = GU\cdot\frac{\sum_{\mu(i,j)\neq 0} v(i,j)}{\sum_{\mu(i,j)\neq 0}\mu(i,j)} = GU\cdot\frac{\sum_{\mu(i,j)\neq 0}\mu(i,j)\gamma_{(i+j)}}{\sum_{\mu(i,j)\neq 0}\mu(i,j)} \qquad (2.15)$$

Finally, a new crisp output of the fuzzy controller at sampling time nT is calculated as

$$u(nT) = u(nT - T) + GU\cdot\Delta u(nT) \qquad (2.16)$$

where GU is the scalar for incremental output and $u(nT - T)$ is the output of the fuzzy controller at sampling time $(n-1)T$.

2.2. *Analytical analysis of structure of the nonlinear fuzzy controller with linear control rules*

Theorem 1. The structure of the nonlinear fuzzy controller with linear control rules is the sum of a global two-dimensional multilevel relay and a local nonlinear PI controller.

Proof. The author first proves the theorem in the situations in which both e^* and r^* are within the interval $[-L, L]$. Others situations will be dealt with later.

(A) Both e^* and r^* are within the interval $[-L, L]$. With losing generality, assume that

$$iS \le e^* \le (i+1)S, \quad jS \le r^* \le (j+1)S \qquad (2.17)$$

Being fuzzified, the memberships of e^* and r^* are obtained as

$$\mu_i(e^*) = -\frac{1}{S}[e^* - (i+1)S], \quad \mu_{i+1}(e^*) = \frac{1}{S}[e^* - iS], \qquad (2.18)$$

$$\mu_j(r^*) = -\frac{1}{S}[r^* - (j+1)S], \quad \mu_{j+1}(r^*) = \frac{1}{S}[r^* - jS], \qquad (2.19)$$

which are the memberships for the members E_i, E_{i+1}, R_j and R_{j+1}, respectively. Membership for all other members of "error" and "rate" is zero. Therefore, only the following four fuzzy control rules are executed.

(r1) If "error" is E_{i+1} and "rate" is R_{j+1} then "output" is $U_{(i+j+2)}$.

(r2) If "error" is E_{i+1} and "rate" is R_j then "output" is $U_{(i+j+1)}$.

(r3) If "error" is E_i and "rate" is R_{j+1} then "output" is $U_{(i+j+1)}$.

(r4) If "error" is E_i and "rate" is R_j then "output" is $U_{(i+j)}$.

Applying equation (2.13) to each of the fuzzy control rules results in the following

(r1*) $v(i+1, j+1) = -\mathrm{Min}\,(\mu_{i+1}(e^*), \mu_{j+1}(r^*))\cdot(i+j+2)V$

(r2*) $v(i+1, j) = -\mathrm{Min}\,(\mu_{i+1}(e^*), \mu_j(r^*))\cdot(i+j+1)V$

(r3*) $v(i, j+1) = -\mathrm{Min}\,(\mu_i(e^*), \mu_{j+1}(r^*))\cdot(i+j+1)V$

(r4*) $v(i, j) = -\mathrm{Min}\,(\mu_i(e^*), \mu_j(r^*))\cdot(i+j)V$

To determine the results of the Min operations in (r1*) to (r4*) the author configures a square by the intervals $[iS, (i+1)S]$ and $[jS, (j+1)S]$ and divides the square into eight regions as shown in Fig. 2. In different regions, $\mu_i(e^*)$, $\mu_{i+1}(e^*)$, $\mu_j(r^*)$ and $\mu_{j+1}(r^*)$ have different relationships in terms of the magnitudes of the memberships. The outcomes of evaluating the Min operations are illustrated in Table 2. Since the fuzzy control rules r2 and r3 generate two

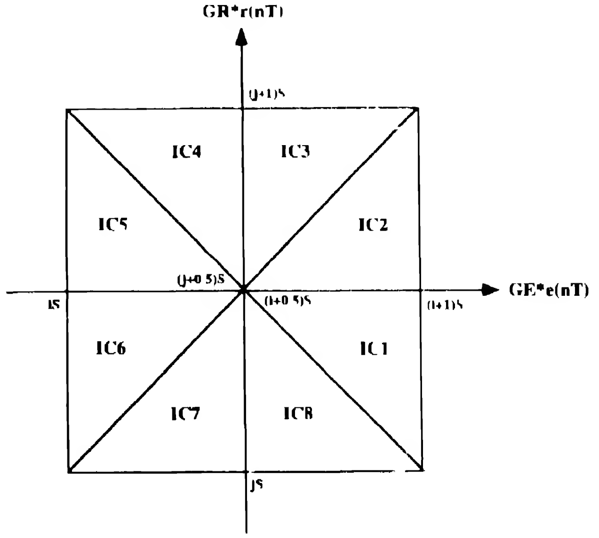


FIG. 2. Possible input combinations (IC) of scaled error, e^*, and scaled rate change of error, r^*, of process output which must be considered to carry out the Min operation in (r1*) to (r4*) when both e^* and r^* are within the interval $[-L, L]$.

TABLE 2. RESULTS OF EVALUATING THE MIN OPERATIONS IN ($r1^*$) TO ($r4^*$) FOR ALL COMBINATIONS OF INPUTS USING ZADEH FUZZY LOGIC AND (MIN) WHEN SCALED ERROR AND RATE CHANGE OF ERROR OF PROCESS OUTPUT ARE WITHIN THE INTERVAL $[-L, L]$. THE INPUT COMBINATIONS OF SCALED ERROR AND RATE CHANGE OF ERROR ARE SHOWN GRAPHICALLY IN FIG 2

Region	r1*	r2*	r3*	r4*
IC1 and IC2	$\mu_{j+1}(r^*)$	$\mu_j(r^*)$	$\mu_i(e^*)$	$\mu_i(e^*)$
IC3 and IC4	$\mu_{i+1}(e^*)$	$\mu_j(r^*)$	$\mu_i(e^*)$	$\mu_j(r^*)$
IC5 and IC6	$\mu_{i+1}(e^*)$	$\mu_{i+1}(e^*)$	$\mu_{j+1}(r^*)$	$\mu_j(r^*)$
IC7 and IC8	$\mu_{j+1}(r^*)$	$\mu_{i+1}(e^*)$	$\mu_{j+1}(r^*)$	$\mu_i(e^*)$

memberships for the same member, $U_{-(i+j+1)}$, the equation (2.14) is needed to calculate the combined membership for $U_{-(i+j+1)}$. For the ICI to IC4 regions,

$$\mu_i(e^*) + \mu_j(r^*) = 1 - \frac{[e^* - (i+0.5)S] + [r^* - (j+0.5)S]}{S} \leq 1, \quad (2.20)$$

because

$$0 \leq [e^* - (i+0.5)S] + [r^* - (j+0.5)S] \leq S. \quad (2.21)$$

Similarly, it is easy to prove

$$\mu_{i+1}(e^*) + \mu_{j+1}(r^*) \leq 1, \text{ for IC5 to IC8 regions.} \quad (2.22)$$

Hence, the combined membership for $U_{-(i+j+1)}$ is always the sum of the memberships being ORed. Replacing the Min operations in ($r1^*$) to ($r4^*$) with their corresponding outcomes in Table 2 and using the defuzzification algorithm (2.15) in connection with (2.7), (2.9), (2.18) and (2.19), the scaled incremental output, $GU \cdot \Delta u(nT)$, for all eight regions can be found as follows:
for IC1, IC2, IC5 and IC6 regions

$$GU \cdot \Delta u(nT) = -(i+j+1)\frac{GU \cdot H}{N-1} - \frac{[GE \cdot e(nT) - (i+0.5)S] + [GR \cdot r(nT) - (j+0.5)S]}{2S - 2|GE \cdot e(nT) - (i+0.5)S|} \times \frac{GU \cdot H}{N-1} \quad (2.23)$$

for IC3, IC4, IC7 and IC8 regions

$$GU \cdot \Delta u(nT) = -(i+j+1)\frac{GU \cdot H}{N-1} - \frac{[GE \cdot e(nT) - (i+0.5)S] + [GR \cdot r(nT) - (j+0.5)S]}{2S - 2|GR \cdot r(nT) - (j+0.5)S|} \times \frac{GU \cdot H}{N-1}. \quad (2.24)$$

$GU \cdot \Delta u(nT)$ consists of two parts. The first part is $-(i+j+1)GU \cdot H/(N-1)$, which is a two-dimensional multilevel relay with respect to i and j. Note that the multilevel relay, denoted as Relay(i, j) can be rewritten as

$$\begin{aligned}\text{Relay}(i,j) &= -(i+j+1)\frac{GU \cdot H}{N-1} \\ &= -((i+0.5)S + (j+0.5)S)\frac{GU \cdot H}{S(N-1)}, \\ &= -((i+0.5)S + (j+0.5)S)\frac{GU \cdot H}{2L}. \quad (2.25)\end{aligned}$$

The point $((i+0.5)S, (j+0.5)S)$ is the coordinate of the center of the square shown in Fig. 2. Evidently, the multilevel relay contributes its control action according to the absolute position, with respect to the entire scaled input state plane, of the center of the square in which the current scaled input state (e^*, r^*) lies. Therefore, the author calls the multilevel relay a "global" multilevel relay. The second part of $GU \cdot \Delta u(nT)$ is a nonlinear nonfuzzy controller, which is denoted as $\delta u(i, j)$. The equations (2.23) and (2.24) indicate that $\delta u(i, j)$ is calculated according to the relative position of the current scaled input state ($GE \cdot e(nT)$, $GR \cdot r(nT)$) with respect to the center of the square, $((i+0.5)S, (j+0.5)S)$, in which the current scaled input state lies. Therefore, one can see that the role of the nonlinear controller is to locally adjust the control action generated by the global multilevel relay. The author calls such a controller a "local" nonlinear controller.

A regular discrete-form linear PI controller whose output becomes zero when its inputs, $e(nT)$ and $r(nT)$, reach a steady-state $((i+0.5)S/GE, (j+0.5)S/GR)$, can be expressed as

$$\delta u_{PI}(i,j) = -\left(K_i\left[e(nT) - \frac{(i+0.5)S}{GE}\right] + K_p\left[r(nT) - \frac{(j+0.5)S}{GR}\right]\right), \quad (2.26)$$

where K_p and K_i are the proportional-gain and integral-gain, respectively. Therefore, the local nonlinear controller is actually a nonlinear PI controller with a local and changing steady-state $((i+0.5)S/GE, (j+0.5)S/GR)$:

$$\delta u(i,j) = -\left(K_i(e^*, r^*)\left[e(nT) - \frac{(i+0.5)S}{GE}\right] + K_p(e^*, r^*)\left[r(nT) - \frac{(j+0.5)S}{GR}\right]\right).$$

The proportional-gain and integral-gain change with input states and are described in equation (3.1)

(B) Either e^* or r^* is outside the interval $[-L, L]$.

To analytically describe the behavior of the nonlinear fuzzy controller when either e^* or r^* is outside the interval $[-L, L]$, the author divides the scaled input state plane outside the square configured by the interval $[-L, L]$ on the scaled error axis and the interval $[-L, L]$ on the scaled rate axis into 12 regions, as shown in Fig. 3. By using the same method described above, $GU \cdot \Delta u(nT)$ can be analytically derived for the regions, as shown in Table 3. According to Table 3, the nonlinear fuzzy controller becomes the sum of a global one-dimensional multilevel relay and a local linear P controller with a local and changing steady-state for the IC9, IC10, IC13 and IC14 regions, and the sum of a global one-dimensional multilevel relay and a local linear integral (I) controller with a local and changing steady-state for the IC11, IC12, IC15 and IC16 regions. The nonlinear fuzzy controller generates its maximum increment ($GU \cdot H$) and decrement ($-GU \cdot H$) in the IC19 and IC17 regions, respectively. For the IC18 and IC20 regions, the increment is zero. ■

It should be noted that when a scaled input state (e^*, r^*) is on a boundary of two adjacent regions, $GU \cdot \Delta u(nT)$ calculated by using the formula of either region is the same. In other words, there is no discrepancy in control action.

3. *Properties of the nonlinear fuzzy controller with linear control rules*

3.1. *Dynamic change of local nonlinear PI controller gains.* Comparing (2.23) and (2.24) with (2.26), one can see that the proportional-gain and integral-gain of the local

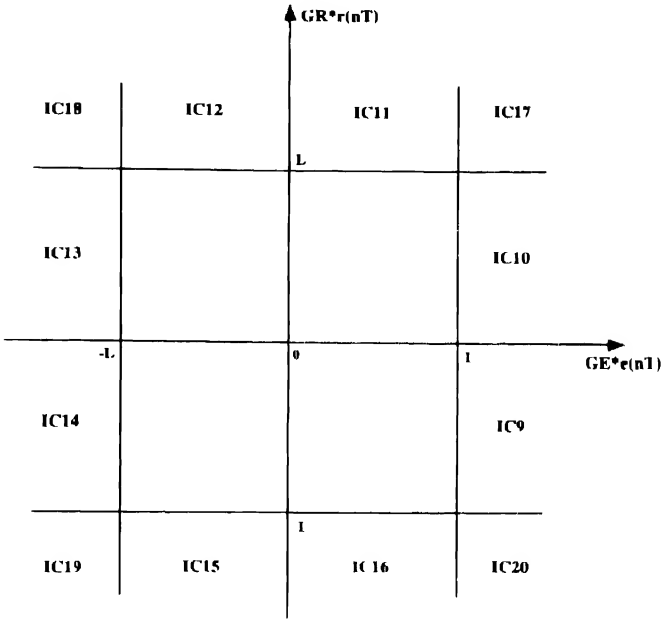


FIG 3 Possible input combinations (IC) of scaled error, e^* and scaled rate change of error, r^*, of process output which must be considered to carry out the Min operation in (r1*) to (r4*) when either e^* or r^* is outside the interval $[-L, L]$

TABLE 3 THE SCALED INCREMENTAL OUTPUT OF THE FUZZY CONTROLLERS $GU \cdot \Delta(nT)$, WHEN EITHER SCALED ERROR $GE \cdot e(nT)$ OR SCALED RATE CHANGE OF ERROR $GR \cdot r(nT)$, OF PROCESS OUTPUT IS OUTSIDE THE INTERVAL $[-L, L]$ THE INPUT COMBINATIONS OF SCALED ERROR AND RATE CHANGE OF ERROR ARE SHOWN GRAPHICALLY IN FIG 2

IC9 and IC10	$-(J+j)\frac{GU \cdot H}{N-1} - (GR \cdot r(nT) - j \cdot S)\frac{GU \cdot H}{2L}$
IC11 and IC12	$(i+J)\frac{GU \cdot H}{N-1} - (GE \cdot e(nT) - i \cdot S)\frac{GU \cdot H}{2L}$
IC13 and IC14	$-(-J+j)\frac{GU \cdot H}{N-1} - (GR \cdot r(nT) - j \cdot S)\frac{GU \cdot H}{2L}$
IC15 and IC16	$-(i-J)\frac{GU \cdot H}{N-1} - (GE \cdot e(nT) - i \cdot S)\frac{GU \cdot H}{2L}$
IC17	$-GU \cdot H$
IC18	0
IC19	$GU \cdot H$
IC20	0

nonlinear PI controller vary with input state and are

$$
\begin{aligned}
K_p(e^*, r^*) &= \frac{GR \cdot GU \cdot H}{2(N-1)S}\beta(e^*, r^*) \\
&= \frac{GR \cdot GU \cdot H}{4L}\beta(e^*, r^*), \qquad (3\ 1)\\
K_i(e^*, r^*) &= \frac{GE \cdot GU \cdot H}{2(N-1)S}\beta(e^*, r^*) \\
&= \frac{GE \cdot GU \cdot H}{4L}\beta(e^*, r^*),
\end{aligned}
$$

where

$$\beta(e^*, r^*) = \frac{S}{S - |GE \cdot e(nT) - (i + 0\ 5)S|},$$

for IC1, IC2, IC5 and IC6 regions,

and (3 2)

$$\beta(e^*, r^*) = \frac{S}{S - |GR \cdot r(nT) - (j + 0\ 5)S|},$$

for IC3, IC4, IC7 and IC8 regions

Obviously, the local nonlinear PI controller can automatically adjust the proportional-gain and integral gain to adapt to different scaled input states The further the current scaled input state ($GE \cdot e(nT)$, $GR \cdot r(nT)$) is from the center of the square $((i + 0\ 5)S, (j + 0\ 5)S)$ the larger the proportional gain and integral gain With the constraints on e^* and r^* specified in (2 17), the range of the value of the nonlinear function $\beta(e^*, r^*)$ is calculated as

$$1 \le \beta(e^*, r^*) \le 2 \qquad (3\ 3)$$

Hence, the ranges of the proportional-gain and integral-gain are

$$
\begin{aligned}
\frac{GR \cdot GU \cdot H}{4L} &\le K_p(e^*, r^*) \le \frac{GR \cdot GU \cdot H}{2L} \\
\frac{GE \cdot GU \cdot H}{4L} &\le K_i(e^*, r^*) \le \frac{GE \cdot GU \cdot H}{2L}
\end{aligned}
\qquad (3\ 4)
$$

3 2 *The role of the global multilevel relay and the local nonlinear PI controller in total control action and degree of nonlinearity* The absolute value of maximum Relay(i, j) is $Relay_{max} = (N-2)GU \cdot H/(N-1)$, which is achieved when

$i = j = J - 1$ or $i = j = -J$. The absolute value of maximum $\delta u(i, j)$ is $\delta u_{\max} = GU \cdot H/(N-1)$, which is achieved when $GE \cdot e(nT) = (i+1)S$ and $GR \cdot r(nT) = (j+1)S$ or when $GE \cdot e(nT) = i \cdot S$ and $GR \cdot r(nT) = j \cdot S$. The author defines the ratio

$$\rho = \frac{\delta u_{\max}}{\text{Relay}_{\max} + \delta u_{\max}} \times 100\% = \frac{1}{N-1} \times 100\%, \quad (3.5)$$

to describe (1) the role of the local nonlinear PI controller and the role of the global multilevel relay in total control action; and (2) the degree of nonlinearity of the nonlinear fuzzy controllers as N changes. The smaller the ratio ρ, the less significant the role of the local nonlinear PI controller in total control action and the more significant the role of the global multilevel relay in total control action. When $N = 3$, ρ reaches its maximum, 50%, which indicates that the local nonlinear PI controller plays as important a role as does the global multilevel relay.

According to (3.4), the ranges of the proportional-gain and integral-gain are independent from N. That means that the ability of the local nonlinear PI controller to adapt locally to input state is the same for fuzzy controllers as N changes. However, it should be noted that the role of the local nonlinear PI controller in total control action is governed by the ratio ρ and therefore is different when N is different.

The ratio ρ also describes the degree of nonlinearity of the nonlinear fuzzy controllers. The smaller the ratio ρ, the finer the resolution of the output of the global multilevel relay and therefore the less nonlinear the fuzzy controller.

3.3. *The structure of the nonlinear fuzzy controllers when* $N \to \infty$.

Theorem 2 (*Limit theorem*). The nonlinear fuzzy controller with linear control rules becomes a linear PI controller as $N \to \infty$.

Proof. When $N \to \infty (J \to \infty)$, the control action from the local nonlinear PI controller approaches zero according to (2.23) and (2.24), that is $\delta u(i, j) \to 0$. On the other hand, the control action from the global multilevel relay becomes

$$\text{Relay}(i, j) = -(i + j + 1)\frac{GU \cdot H}{2J} = -\frac{GU \cdot H}{2}\left(\frac{i}{J} + \frac{j}{J}\right). \quad (3.6)$$

Because

$$iS = i\frac{L}{J}, \quad (i+1)S = (i+1)\frac{L}{J},$$
$$jS = j\frac{L}{J}, \quad (j+1)S = (j+1)\frac{L}{J}, \quad (3.7)$$

the inequalities (2.17) can be written as

$$\frac{i}{J} \le \frac{e^*}{L} \le \frac{i+1}{J} \quad \text{and} \quad \frac{j}{J} \le \frac{r^*}{L} \le \frac{j+1}{J} \quad (3.8)$$

and hence

$$\frac{e^*}{L} \to \frac{i}{J} \quad \text{and} \quad \frac{r^*}{L} \to \frac{j}{J}, \quad (3.9)$$

when $N \to \infty$ (therefore $J \to \infty$, $i \to \infty$ and $j \to \infty$). Substituting (3.9) into (3.6), yield

$$\text{Relay}(i, j) = -\frac{GU \cdot H}{2L}(e^* + r^*)$$
$$= -\left(\frac{GU \cdot H \cdot GE}{2L}e(nT) + \frac{GU \cdot H \cdot GR}{2L}r(nT)\right). \quad (3.10)$$

and hence

$$GU \cdot \Delta u(nT) = \text{Relay}(i, j)$$
$$\left(\frac{GU \cdot H \cdot GE}{2L}e(nT) + \frac{GU \cdot H \cdot GR}{2L}r(nT)\right). \quad (3.11)$$

Therefore, one can immediately conclude that the nonlinear fuzzy controller (and the global multilevel relay) becomes a regular linear PI controller when N is ∞ (ρ is zero). The corresponding K_p and K_i are

$$K_p = \frac{GR \cdot GU \cdot H}{2L} \qquad K_i = \frac{GE \cdot GU \cdot H}{2L}. \quad (3.12)$$

4. *Relationship between fuzzy controllers with* $N \ge 3$ *and that with* $N = 2$

In Ying *et al.* (1990), we analytically proved that a simplest possible ($N = 2$) nonlinear fuzzy controller constructed in the same way as those in this paper was a nonlinear PI controller:
for IC1, IC2, IC5 and IC6 regions

$$du(NT) = -\frac{GE \cdot e(nT) + GR \cdot r(nT)}{2L - GE \cdot |e(nT)|}\frac{GU \cdot H}{2}, \quad (4.1)$$

or for IC3, IC4, IC7 and IC8 regions

$$du(nT) = -\frac{GE \cdot e(nT) + GR \cdot r(nT)}{2L - GR \cdot |r(nT)|}\frac{GU \cdot H}{2}, \quad (4.2)$$

where

$$GE \cdot |e(nT)| \le L \quad \text{and} \quad GR \cdot |r(nT)| \le L. \quad (4.3)$$

Based on (4.1) and (4.2), the simplest possible nonlinear fuzzy controller does not include the multilevel relay and hence the nonlinear PI controller is a global controller having one single global and fixed steady-state (0, 0). It can be easily proven that the fuzzy controller with $N = 2$ can be expressed as

$$du(nT) = -[K_i(e^*, r^*)e(nT) + K_p(e^*, r^*)r(nT)], \quad (4.4)$$

where the gains and the ranges of the gains are the same as those in (3.1) and (3.4), which indicates that the local nonlinear PI controller can adjust its gains to adapt to different input states as much as the global nonlinear PI controller can. However, there is a fundamental difference between these two nonlinear PI controllers. That is, the role the local nonlinear PI controller can play in total control action is less significant because the role is governed by the ratio ρ. The role is small when N is large. On the other hand, the global nonlinear PI controller contributes sole control action and therefore its ratio ρ is 100% by definition, which also means the fuzzy controller with $N = 2$ is more nonlinear than any other fuzzy controllers.

5. *Conclusions*

The author concludes that the nonlinear fuzzy controllers with $N \ge 3$ consist of a global multilevel relay and a local nonlinear PI controller similar to the global nonlinear PI controller when $N = 2$ but with a local and changing steady-state $((i + 0.5)S/GE,\ (j + 0.5)S/GR)$. The consequences of employing more than two members ($N \ge 3$) for input fuzzy sets are (1) introducing the global multilevel relay with resolution $GU \cdot H/(N-1)$; and (2) reducing the role of the local nonlinear PI controller in total control action from 100% to $1/(N-1)$. Larger N makes the fuzzy controllers less nonlinear. As N approaches ∞, the nonlinearity disappears and the fuzzy controller becomes a linear PI controller. The degree of nonlinearity of the fuzzy controllers and the role of the local nonlinear PI controllers in total control action are quantitatively described by introducing the ratio ρ. The fuzzy controller with $N = 2$, whose ρ is 100%, is the most nonlinear fuzzy controller.

Acknowledgements—The author is grateful to Prof. Louis C. Sheppard for his support in the research. Sincere thanks also go to Ms Karin Elder for her proofreading of the manuscript and to the referees for their helpful comments and suggestions about the manuscript.

References

Buckley, J. J. (1989a). Further results for the linear fuzzy controller. *Kybernetes*, **18**, 48–55.
Buckley, J. J. (1989b). Nonlinear fuzzy controller. *Information Sciences*, to appear.
Buckley, J. J. and H. Ying (1989). Fuzzy controller theory:

limit theorems for linear fuzzy control rules. *Automatica*, **25**, 469–472.

Buckley J. J. and H. Ying (1990). Linear fuzzy controller: it is a linear nonfuzzy controller. *Information Sciences*, **51**, 183–192.

Buckley, J. J. (1990). Further limit theorems for linear control rules. *Fuzzy Sets and Systems*, **36**, 225–233.

Siler, W. and H. Ying (1989). Fuzzy control theory: the linear case. *Fuzzy Sets and Systems*, **33**, 275–290.

Wang, P. Z., H. M. Zhang and W. Xu (1990). Pad-analysis of fuzzy control stability. *Fuzzy Sets and Systems*, **38**, 27–42.

Ying, H. (1987). Fuzzy control theory. Technical Report. Department of Biomedical Engineering, University of Alabama at Birmingham.

Ying, H., W. Siler and J. J. Buckley (1988). Fuzzy control theory: a nonlinear case. *Proc. of NASA's Conference on Artificial Neural Systems and Fuzzy Logic*, Houston, U.S.A., 2–3 May.

Ying, H., W. Siler and J. J. Buckley (1990). Fuzzy control theory: a nonlinear case. *Automatica*, **26**, 513–520.

Automatica, Vol. 29, No. 2, pp. 507–509, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Brief Paper

Polynomial LQ Optimization for the Standard Control Structure: Scalar Solution*

K. J. HUNT† and M. ŠEBEK‡

Key Words—Optimal control; algebraic system theory; polynomials

Abstract—The problem of LQ optimization for the "standard" control structure is studied. We obtain a solution for the scalar case of this problem using the polynomial equation approach. In solving the optimal regulation problem using polynomial techniques we derive a couple of linear polynomial equations which together define the unique optimal regulator. The coefficients of these equations are obtained from polynomial spectral factorization.

1. *Introduction*

IN THIS WORK we focus on the synthesis of LQ-optimal control systems using the polynomial equation approach (Kučera, 1979). In particular, we derive the optimizing solution for the scalar case of the Standard Control Structure. This polynomial approach provides a distinctive alternative to the state-space (Kwakernaak and Sivan, 1972) and Wiener–Hopf (Youla *et al.*, 1976) optimization procedures.

The standard structure (Fig. 1) has been used as the basis of H_∞ optimal control problems (Francis, 1987; Kwakernaak, 1990). It has also been studied in H_2-optimal control problems; firstly using the Wiener–Hopf approach (Park and Bongiorno, 1989) and then the state-space approach (Doyle *et al.*, 1989).

In the multivariable case all sensor outputs are incorporated into y, while the regulated variable is the vector z. The vector ζ denotes all exogenous inputs to the system including, for example, disturbances and references. The vector u represents the control inputs. In the paper we restrict our attention to the case where all these signals are scalar. This provides an important first step towards solution of the multivariable problem, which is currently under study (see Hunt *et al.*, 1991).

The stability of the scalar standard structure was fully analysed in Kučera (1986). Under the assumption of full internal stability of the overall system a solution for the LQ optimization problem for the standard structure was obtained by Hunt *et al.* (1992). The results presented in that work are closely related to the solution given here. Here, however, we recognise that the standard structure is an artificial construction incorporating a variety of control problems and therefore require only stability of the feedback part of the system.

A preliminary version of this paper was published by Hunt and Šebek (1991). The solution of a closely related problem may be found in Hunt and Kučera (1992).

1.1. *Notation* All systems considered in this work are assumed to be linear, time-invariant and discrete-time. The systems are described by means of real polynomials in one indeterminate d. Where d is to be interpreted as the unit delay operator described by the relation $dx(t) = x(t-1)$ for any sequence $x(t)$. The reader is referred to the work by Kučera (1979) for details.

For simplicity, the arguments of polynomials are usually omitted; a polynomial $X(d)$ is denoted by X. For any polynomial $X(d)$ we denote its adjoint in negative powers of d by $X^*(d)$. For any polynomial $X(d)$ we define $\langle X \rangle$ as the constant term independent of d. Stable polynomials are those with all their zeros having magnitude greater than unity.

2. *Problem formulation*

2.1. *Plant*. The plant under consideration is governed by the polynomial model

$$A_y y = B_y u + E_y \zeta, \tag{1}$$

$$A_z z = C_z u + D_z \zeta, \tag{2}$$

where y is the measurement, z is the variable to be regulated, u is the control input, and ζ is a stochastic disturbance. A_y, B_y, E_y and A_z, C_z, D_z are two triples of coprime polynomials.

Further, we write A_y and A_z in terms of their stable/unstable factorizations as follows.

$$A_y = A_y^+ A_y^-, \tag{3}$$

$$A_z = A_z^+ A_z^-, \tag{4}$$

where $^+$ denotes a stable polynomial and $^-$ denotes an unstable polynomial.

Moreover, it is assumed that.

(A.1) the disturbance ζ is a stationary, zero-mean white noise sequence with intensity $\phi > 0$;
(A.2) the transfer function $A_y^{-1}B_y$ is strictly causal so that $\langle B_y \rangle = 0$ and $\langle A_y \rangle = 1$;
(A.3) the unstable part of A_z divides A_y, i.e. there exists some polynomial K such that $A_z^- K = A_y^-$

Remarks

(1) The plant description (1)–(2) is quite general and corresponds to a plant described by a transfer matrix where the denominator of each entry equals the least common denominator of the corresponding row.
(2) Assumption (A.3) is a standard assumption (Park and Bongiorno, 1989; Doyle *et al.*, 1989) meaning that all unstable modes appearing in z must also be seen in y to be controlled.

2.2. *Regulator*. The linear regulator which operates on the plant measurement y is described by

$$u = -Ry. \tag{5}$$

* Received 30 November 1990; revised 24 September 1991, revised 9 June 1992; received in final form 8 August 1992. The original version of this paper was presented at the IFAC Symposium on Design Methods of Control Systems which was held in Zurich, Switzerland during September 1991. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor R. V. Patel under the direction of Editor H. Kwakernaak.

† Control Group, Department of Mechanical Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

‡ Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Pod vodárenskou věží 4, 182 08 Praha 8, Czechoslovakia. Paper written while on leave at the Faculty of Applied Mathematics, University of Twente, The Netherlands.

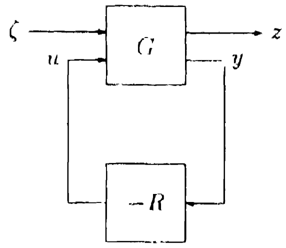


FIG. 1. Standard feedback system.

The transfer function R is written in rational form as

$$R = X^{-1}Y, \tag{6}$$

where X and Y are polynomials, and $\langle X \rangle = 1$. Thus, we seek a regulator described by (5)–(6) and are only interested in those with $X(d)$ invertible.

We also write the regulator transfer function in the form

$$R = p^{-1}q, \tag{7}$$

where p and q are rational functions to be further specified below.

A deep analysis of the internal stability of feedback systems has been given by Kučera (1979, 1986). This analysis shows that to ensure stability of the feedback part of the system we must have:

(1) p and q must be stable rational functions which satisfy the relation

$$pA_y + qB_y = 1. \tag{8}$$

(2) R is realised without unstable hidden modes.

We return to the relation (8) in the sequel when considering questions of stability. To satisfy the remaining condition we proceed under the assumption:

(A.4) The regulator is free of unstable hidden modes.

Note here that in considering stabilizability of the system we require internal stability of only the feedback part of the system (later, when aiming to achieve a finite cost we also require stability of the closed-loop transfer function between ζ and z). This concept recognises that the standard control structure is an artificial construct incorporating a variety of control problems, some of which lead naturally to a realization in the standard form which is not completely internally stable. In our previous solution to the LQ optimization problem for the standard structure we required complete stability of the overall system (Hunt *et al.*, 1992). In our approach presented here the solvability conditions are relaxed.

2.3 *Cost.* The desired optimal regulator evolves from minimization of the cost function

$$J = \langle R_c\phi_u + Q_c\phi_z \rangle, \tag{9}$$

where ϕ_u and ϕ_z are the correlation functions of u and z in the steady-state, respectively. R_c and Q_c are real, non-negative weights, not simultaneously zero.

The optimal regulator problem is to minimize the cost (9) subject to the constraint that the feedback part of the system be internally stable.

3. *Problem solution*

We define the spectral factors D_c and D_f by the relations

$$A_z^*R_cA_z + C_z^*Q_cC_z = D_c^*D_c, \tag{10}$$

$$E_y\phi E_y^* = D_fD_f^*, \tag{11}$$

and note that we are interested in solutions D_c, D_f which are stable polynomials. It is clear that necessary and sufficient conditions for stable spectral factors to exist are that:

(1) R_cA_z and Q_cC_z have no common factors with zeros on the unit circle.

(2) E_y has no zeros on the unit circle.

For simplicity, we proceed here under the assumption that the given data make the problem regular (that the above conditions hold), i.e. that:

(A.5) there exist stable solutions D_c and D_f to equations (10)–(11).

We may now summarize the solution to the optimal regulation problem:

Theorem 1. The optimal regulation problem using measurement feedback has a solution if and only if:

(C.1) the greatest common divisor of A_y and B_y is a stable polynomial.

(C.2) there exists some polynomial H such that $A_zH = E_yC_z - D_zB_y$.

The optimal regulator is a realization of the unique transfer function

$$R = P^{-1}QA_y^{+}. \tag{12}$$

Here, P and Q is the unique solution having the property $\langle Z \rangle = 0$ of the linear polynomial equations

$$D_c^*D_f^*Q - Z^*A_z = L_1, \tag{13}$$

$$D_c^*D_f^*KP + Z^*A_z^{\dagger}B_y = L_2, \tag{14}$$

where the polynomials L_1 and L_2 satisfy

$$L_1 = C_z^*Q_cE_y^*\phi D_z, \tag{15}$$

$$L_2 = D_fD_f^*A_z^*R_cA_z^{\dagger} + E_y^*C_z^*Q_c\phi H. \tag{16}$$

Proof.

The control and measurement sequences may be written as

$$u = -qE_y\zeta, \tag{17}$$

$$z = A_z^{-1}(D_z - qC_zE_y)\zeta. \tag{18}$$

The corresponding correlation functions are

$$\phi_u = qE_y\phi E_y^*q^*, \tag{19}$$

$$\phi_z = A_z^{-1}(D_z - qC_zE_y)\phi(D_z - qC_zE_y)^*A_z^{-1*} \tag{20}$$

Introducing the spectral factorization (11) ϕ_u becomes

$$\phi_u = qD_fD_f^*q^* \tag{21}$$

The proof now proceeds in three parts. In part (a) we determine the optimizing regulator, in part (b) we examine the stability of the closed-loop transfer function in order to determine whether the cost is finite, and in part (c) we show that the feedback part of the system is internally stable.

(a) *Optimization*

The cost function (9) may now be written as

$$J = \langle R_cqD_fD_f^*q^* + Q_cA_z^{-1}(C_zqD_fD_f^*q^*C_z^* - q^*C_z^*E_y^*D_z\phi - qC_zE_yD_z^*\phi + D_zD_z^*\phi)A_z^{-1*} \rangle. \tag{22}$$

Introducing the spectral factorization (10) this equation becomes

$$J = \langle A_z^{-1}qD_fD_cD_c^*D_f^*q^*A_z^{-1*} - Q_cA_z^{-1}(q^*C_z^*E_y^*D_z\phi + qC_zE_yD_z^*\phi - D_zD_z^*\phi)A_z^{-1*} \rangle. \tag{23}$$

After completing the squares, and some algebraic manipulation, the cost becomes

$$J = \langle V_1^*V_1 \rangle + \langle V_2 \rangle, \tag{24}$$

where

$$V_1 = A_z^{-1}\left(D_cD_fq - \frac{C_z^*E_y^*D_z\phi Q_c}{D_c^*D_f^*}\right), \tag{25}$$

$$V_2 = \frac{\phi Q_cD_zD_z^*R_c}{D_cD_c^*}. \tag{26}$$

We note that only the term V_1 depends on the regulator; V_2 does not include q. We therefore proceed by attempting to minimize the term V_1.

The second term in V_1 contains both causal and non-causal parts. These parts may be split by introducing the

polynomials Q and Z as follows

$$\frac{C^*E_2^*D\phi Q}{A\,D^*D_f^*} = \frac{Q}{A} - \frac{Z^*}{D^*D_f^*} \tag{27}$$

Note that this equation is just the linear polynomial equation (13). V_1 now reads

$$V_1 = \frac{D_c D_f q}{A} - \frac{Q}{A} + \frac{Z^*}{D^*D_f^*} \tag{28}$$

Under the constraint $\langle V\rangle = \langle V^*\rangle = 0$ the contribution of the final term in (28) to the cost vanishes. To minimize the cost we therefore set the causal part of V_1 to zero

$$\frac{D\,D_f q}{A} - \frac{Q}{A} = 0 \tag{29}$$

or

$$q = \frac{Q}{D\,D_f} \tag{30}$$

From the relation (8) the remaining regulator term is

$$p = A_s^{-1}(1 - qB_s) \tag{31}$$

and in (30) and (31), together with (7) we have the regulator which minimizes the cost. Moreover, due to the assumption of the existence of stable spectral factors D and D_f (A 5), q defined by (30) is seen to be stable. However, it is not clear from (31) that p is a stable rational function. To show that the feedback part of the system is stable we must still show that p is stable.

(b) *Finite cost*

The cost function will be finite if and only if all the rational functions in equations (17) and (18) are stable. Stability of q has already been proven so we must still deal with

$$z = A_s^{-1}(D - q(\quad E\quad))\varepsilon \tag{32}$$

Clearly, proving that the optimal cost is finite amounts to showing that the transfer function between and ε (the closed loop transfer function) is stable.

It is straightforward to rewrite the above expression in the form

$$A_s^{-1}(D\,A\,p - (C\,I - D\,B\,)q)\xi \tag{33}$$

Here q has already been shown to be a stable rational function. We postpone discussion of the stability of p until part (c) below. Then taking account of the Assumption (A 3), it is clear that the remaining necessary and sufficient condition for stability of the closed loop transfer function (and therefore for finiteness of the cost) is

(1) The unstable part of A must divide $(C\,I - D\,B\,)$, i.e. there must exist some polynomial H such that $A\,H = C\,L - D\,B_s$. This is Condition (C 2) in the theorem.

We note finally that in arriving at the above condition any possible cancellations between the unstable part of A and q in (33) were ruled out due to the constraints (A 3) and (8).

(c) *Internal stability*

We now investigate the stability of the rational function p. From equation (31) p is given as

$$p = (1 - B_s q)A_s^{-1} \tag{34}$$

Substituting firstly for q from (30) and then for Q from (13), p may be written after some rearrangement as

$$p = (L_s^*C^*\phi Q\,A\,H + L_s\phi E_s^*A\,R\,A^* - Z^*A\,B_s) \times (D\,D_f D^* D_f^* A\,)^{-1} \tag{35}$$

The denominator of this expression includes the unstable factors $D^*D_f^*$ and A_s. In order to obtain a stable p the numerator must include these unstable terms as factors, i.e. there must exist some polynomial P such that the numerator of the above expression is equal to $D^*D_f^*A_s P$. Substituting this value into the numerator we obtain

$$p = \frac{P}{D\,D_f A} \tag{36}$$

Also from the substitution we obtain

$$A_s D^*D_f^*P = E_s^*C_2^*\phi Q\,A\,H + L_s\phi E_s^*A\,R\,A^* - Z^*A\,B_s \tag{37}$$

After substituting $A_s = A\,K$ and cancelling the A term throughout, this is seen to be the linear equation (14).

Moreover, it is now clear from (36) that p is stable due to the definition of the spectral factors D and D_f (and Assumption A 5) and the feedback system is seen to be stable. Assumption (A 4) ensures that no unstable hidden modes can destroy stability of the system. Putting together equations (30) and (36) with equation (7) the optimal regulator transfer function (12) results.

Finally we note that Condition (C 1) in Theorem 1, as well as ensuring stabilizability of the plant, also ensures the existence of a unique solution to the pair of equations (13), (14) with $\langle V\rangle = 0$ (see Kučera (1979) for details).

Corollary 1. The polynomials P and Q equations (13)–(14) also satisfy the linear polynomial equation

$$A\,P + B\,Q = D\,D_f \tag{38}$$

Proof. Substitute for p and q from equations (30) and (36) into equation (8) (the result can also be proved by eliminating the coupling term Z from the equations (13), (14)).

4. *Conclusions*

A solution to the optimal regulation problem for the standard control structure has been given. We considered a scalar external polynomial model of the plant. The optimal regulator polynomials were shown to satisfy a pair of coupled linear polynomial equations. The coefficients of these equations are obtained by two spectral factorizations. It was shown that one of the polynomial equations guaranteed optimality and the other stability of the feedback part of the system.

Acknowledgements — K. J. Hunt is currently a Personal Research Fellow of the Royal Society of Edinburgh. The authors acknowledge the contribution of colleague Vladimir Kučera through many discussions surrounding this line of work.

References

Doyle, J. C., K. Glover, P. P. Khargonekar and B. A. Francis (1989). State space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Aut. Control*, **34**, 831–847.

Francis, B. A. (1987). *A Course in H_∞ Control Theory*. Springer Verlag, Berlin.

Hunt, K. J. and V. Kučera (1992). The standard H_2 optimal control problem: a polynomial solution. *Int. J. Control*, Volume 56, pp 245–251.

Hunt, K. J. and M. Šebek (1991). Polynomial solution to the standard LQ optimal control problem: scalar case. *Proc. IFAC Symposium on Design Methods for Control Systems*, Zurich, Switzerland.

Hunt, K. J., M. Šebek and V. Kučera (1991). Polynomial approach to H_2 optimal control: the multivariable standard problem. *Proc. IEEE Conf. on Decision and Control*, Brighton, U.K.

Hunt, K. J., V. Kučera and M. Šebek (1992). Optimal regulation using measurement feedback: A polynomial approach. *Trans. IEEE Aut. Control*, **37**, 682–685.

Kučera, V. (1979). *Discrete Linear Control: The Polynomial Equation Approach*. Wiley, Chichester.

Kučera, V. (1986). Internal properness and stability in linear systems. *Kybernetika*, **22**, 1–18.

Kwakernaak, H. and R. Sivan (1972). *Linear Optimal Control Systems*. Wiley, New York.

Kwakernaak, H. (1990). The polynomial approach to H_∞ optimal regulation. *Proc. CIME Course on Recent Developments in H_∞ Control Theory*, Como, Italy.

Park, K. and J. J. Bongiorno (1989). A general theory for the Wiener–Hopf design of multivariable control systems. *IEEE Trans. Aut. Control*, **34**, 619–626.

Youla, D. C., H. A. Jabr and J. J. Bongiorno (1976). Modern Wiener–Hopf design of optimal controllers—Part 2: the multivariable case. *IEEE Trans. Aut. Control*, **AC-21**, 319–338.

Automatica, Vol. 29, No. 2, pp 511–516, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Decentralized Estimation and Control with Overlapping Input, State, and Output Decomposition*

ALTUĞ İFTAR†‡

Key Words—Large-scale systems; control theory; estimation theory; decentralized control; decentralized estimation; overlapping decompositions; linear systems.

Abstract—The extension principle, which was introduced recently, is generalized to the case of input, state, and output expansion. Estimator and controller design problems are considered within the framework of the extension principle. Decentralized estimator and controller design with overlapping decompositions is also discussed within the same framework. It is shown that if the extension principle is used then any estimator or controller designed in the expanded spaces is contractible to the original spaces for implementation. Furthermore, it is shown that if an estimator designed for the expanded system achieves good estimation and/or asymptotic estimation then the contracted estimator also achieves good estimation and/or asymptotic estimation for the original system. Similarly, it is also shown that if a controller designed for the expanded system achieves stability and/or good performance then the contracted controller achieves stability and/or good performance for the original system.

1. *Introduction*

MANY OF TODAY'S technological and social problems involve systems which are so complex that it is very costly, if not impossible, to handle such large dimensional systems as a whole. For the purposes of estimation and control, it is usually necessary to decompose the given system into a number of interconnected subsystems. Once a decomposition is available, each subsystem is to be considered independently and the solutions to the subproblems are to be combined to obtain a solution for the original problem.

In order to obtain a useful decomposition, it is essential to identify the parts of a system that are weakly interconnected. However, many large scale systems (e.g. see Özgüner *et al.*, 1988) may consist of subsystems which are strongly connected through certain dynamics (the overlapping part), but weakly connected otherwise. For those systems, disjoint decompositions may easily fail to produce useful results. However, it has been demonstrated that the recently introduced overlapping decompositions (Ikeda and Šiljak, 1980) may produce useful solutions in such cases (e.g. see İftar and Özgüner, 1987).

The estimator or controller design approach within the framework of the inclusion principle (Ikeda *et al.*, 1984) starts with expanding certain spaces (e.g. state, input, and/or output) of a dynamic system with overlapping subsystems into larger spaces such that the subsystems appear as disjoint. Decentralized estimators or decentralized controllers are then designed in the expanded spaces and are finally contracted to the smaller spaces for implementation on the original system.

The earlier results on the inclusion principle, however, were restricted to the expansions and contractions of the state space only. Expansions and contractions of the input and output spaces were first considered by Ikeda and Šiljak (1986) and by Ohta *et al.* (1986). Later İftar and Özgüner (1990) considered controller design with state and input inclusion, and introduced a special case of inclusion called extension. It was shown that if the extension approach is used then any control law designed in the expanded spaces is contractible to the original spaces for implementation. Contractibility is required in order to preserve the desired relations between the expanded and the original systems following the application of appropriate controllers or estimators.

In the present paper, the extension principle is generalized to the case where the output space is also expanded besides the state and the input spaces. Estimator and controller design problems are considered within the framework of the extension principle. Decentralized estimator and controller design with overlapping decompositions is also discussed within the same framework. It is shown that if the extension principle is used then any estimator or controller designed in the expanded spaces is contractible to the original spaces for implementation. Note that, this property may not hold if some other form of inclusion is employed (e.g. see Ikeda and Šiljak, 1986). Furthermore, it is shown that if an estimator designed for the expanded system achieves good estimation and/or asymptotic estimation then the contracted estimator also achieves good estimation and/or asymptotic estimation, respectively, for the original system. Similarly, it is shown that if a controller designed for the expanded system achieves stability and/or good performance then the contracted controller also achieves stability and/or good performance for the original system.

Throughout the paper, $\mathbb{R}^k$ denotes the k-dimensional real vector space, I_k denotes the identity operator on $\mathbb{R}^k$, $\mathbb{R}^{m\times n}$ denotes the space of $m \times n$ real matrices, S^n denotes the space of $n \times n$ symmetric positive semi-definite real matrices, for $x \in \mathbb{R}^n$ and $W \in S^n$, $\|x\|_W$ denotes the weighted semi-norm x^TWx, and $(\cdot)^T$ denotes the transpose of $(\,)$.

2. *Extension principle*

In this section a special case of inclusion, called extension, is introduced. Consider the following linear time-invariant

* Received 14 July 1990; revised 31 October 1991; received in final form 13 July, 1992; the original version of this paper was presented at the 11th IFAC World Congress on Automatic Control at the Service of Mankind which was held in Tallinn, Estonia during August, 1990. The published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor M. Ikeda under the direction of Editor A. P. Sage.

† Department of Electrical and Electronics Engineering, Anadolu Üniversitesi, Bademlik, Eskişehir, Turkey. e-mail: d328@tranavm1.bitnet.

‡ The author was with the Department of Electrical Engineering, University of Toronto, Toronto, Canada during the course of this research.

(LTI) systems:

$$\Sigma: \begin{array}{l} \dot{x} = Ax + Bu, \\ y = Cx, \end{array} \tag{1}$$

and

$$\tilde{\Sigma}: \quad \dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\tilde{u}, \tag{2}$$

Here $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^l$ are, respectively, state, input, and output vectors of the system Σ. Similarly, $\tilde{x} \in \mathbb{R}^{\tilde{n}}$, $\tilde{u} \in \mathbb{R}^{\tilde{m}}$, and $\tilde{y} \in \mathbb{R}^{\tilde{l}}$, are state, input, and output vectors of the system $\tilde{\Sigma}$. The outputs y and $\tilde{y}$ are assumed to be measurable. It is also assumed that $\tilde{n} \geq n$, $\tilde{m} \geq m$, and $\tilde{l} \geq l$. In the sequel, the system Σ is referred to as the original system and $\tilde{\Sigma}$ is referred to as the expanded system. The state, input, and output spaces $\mathbb{R}^n$, $\mathbb{R}^m$, and $\mathbb{R}^l$ of Σ are called, respectively, *original* state, input, and output spaces; similarly, the spaces $\mathbb{R}^{\tilde{n}}$, $\mathbb{R}^{\tilde{m}}$, and $\mathbb{R}^{\tilde{l}}$ of $\tilde{\Sigma}$ are called expanded state, input, and output spaces.

Consider the transformations:

$$T: \mathbb{R}^n \to \mathbb{R}^{\tilde{n}}, \quad \text{rank}\,(T) = n, \tag{3a}$$

$$R: \mathbb{R}^{\tilde{m}} \to \mathbb{R}^m, \quad \text{rank}\,(R) = m, \tag{3b}$$

$$S: \mathbb{R}^l \to \mathbb{R}^{\tilde{l}}, \quad \text{rank}\,(S) = l, \tag{3c}$$

$$T^{\#}: \mathbb{R}^{\tilde{n}} \to \mathbb{R}^n, \quad T^{\#}T = I_n, \tag{3d}$$

and

$$R^{\#}: \mathbb{R}^m \to \mathbb{R}^{\tilde{m}}, \quad RR^{\#} = I_m. \tag{3e}$$

Definition 1. The system $\tilde{\Sigma}$ is an extension of the system Σ, and Σ is a disextension of $\tilde{\Sigma}$, if there exist transformations as in (3a)–(3c) such that for any initial state $x_0 \in \mathbb{R}^n$ of the system Σ and any input $\tilde{u}(t) \in \mathbb{R}^{\tilde{m}}$, $0 \leq t < \infty$, of the system $\tilde{\Sigma}$ the choice

$$\tilde{x}_0 = Tx_0, \tag{4a}$$

and

$$u(t) = R\tilde{u}(t), \quad \forall t \geq 0, \tag{4b}$$

implies that

$$\tilde{x}(t; \tilde{x}_0, \tilde{u}) = Tx(t; x_0, u), \quad \forall t \geq 0, \tag{5a}$$

and

$$\tilde{y}(t; \tilde{x}) = Sy(t; x), \quad \forall t \geq 0. \tag{5b}$$

Remark 1. The extension defined above is a generalization of the extension first defined by İftar and Özgüner (1990) to the case where the output space is also expanded besides the state and the input spaces. It is a special case of inclusion defined by Ikeda and Šiljak (1986). In fact, it is a generalization of unrestriction, first discussed by Ikeda *et al* (1984), to the case of state, input, and output inclusion. However, it is different than the unrestriction defined in Ikeda and Šiljak (1986), where unrestriction was defined for an arbitrary input $u(t)$ in the original input space, and the input in the expanded space is obtained by a transformation: $\tilde{u}(t) = R^{\#}u(t)$. Here, on the other hand, the extension is defined for an arbitrary input $\tilde{u}(t)$ in the expanded input space, and the input in the original space is obtained by the transformation given in (4b). Therefore, for the unrestriction the allowable set of inputs for $\tilde{\Sigma}$ at any time is only an m-dimensional subset of $\mathbb{R}^{\tilde{m}}$, but for the extension it is $\mathbb{R}^{\tilde{m}}$.

The necessary and sufficient conditions for the extension are provided by the following theorem:

Theorem 1. The system $\tilde{\Sigma}$ is an extension of the system Σ if and only if there exist transformations as in (3a)–(3c) such that

$$TA = \tilde{A}T, \tag{6a}$$

$$TBR = \tilde{B}, \tag{6b}$$

and

$$SC = \tilde{C}T. \tag{6c}$$

Proof. Given in the Appendix. □

Next we consider the outputs:

$$w = Dx \in \mathbb{R}^k, \tag{7}$$

of the system Σ and

$$\tilde{w} = \tilde{D}\tilde{x} \in \mathbb{R}^{\tilde{k}}, \tag{8}$$

of the system $\tilde{\Sigma}$. These outputs are not necessarily measurable, but it is assumed that estimates of these quantities are desired. It is also assumed that $k \leq n$, $\tilde{k} \leq \tilde{n}$, and $\tilde{k} \geq k$. In addition to the transformations (3a)–(3c), we define the following transformation:

$$Q: \mathbb{R}^{\tilde{k}} \to \quad \text{rank}\,(Q) = k. \tag{9}$$

We now introduce the inclusion principle for the outputs to be estimated:

Definition 2. The output (8) of the system $\tilde{\Sigma}$ includes the output (7) of the system Σ if there exists transformations as in (3a), (3b), and (9) such that for any initial state $x_0 \in \mathbb{R}^n$ of the system Σ and any input $\tilde{u}(t) \in \mathbb{R}^{\tilde{m}}$, $0 \leq t < \infty$, of the system $\tilde{\Sigma}$ the choice (4a)–(4b) implies that

$$w(t; x) = Q\tilde{w}(t; \tilde{x}), \quad \forall t \geq 0. \tag{10}$$

The necessary and sufficient conditions for the inclusion of outputs to be estimated are provided by the following theorem:

Theorem 2. The output (8) of the system $\tilde{\Sigma}$ includes the output (7) of the system Σ if and only if there exist transformations as in (3a), (3b), and (9) such that

$$DA^i = Q\tilde{D}\tilde{A}^iT,$$

and

$$DA^iBR = Q\tilde{D}\tilde{A}^i\tilde{B}, \tag{11b}$$

for all $i \in \{0, 1, 2, \ldots\}$

Proof The proof follows similar lines to the proof of Theorem 1. □

If the expanded system is an extension of the original system, then a simpler condition may be found.

Corollary 1 Given that the system $\tilde{\Sigma}$ is an extension of the system Σ, the output (8) of the system $\tilde{\Sigma}$ includes the output (7) of the system Σ if there exists a transformation as in (9) such that

$$D = Q\tilde{D}T, \tag{12}$$

where T is the transformation satisfying (6a)

Proof. For $i = 0$ (11a) holds if and only if (12) holds. Given that (6a) holds, (11a) holds $\forall i \in \{0, 1, 2, \ldots\}$ if (12) holds, given that (6a)–(6b) hold, (11b) holds $\forall i \in \{0, 1, 2, \ldots\}$ if (12) holds. Hence, the result follows. □

3. *Estimator and controller design with extension*

In this section, we discuss the design of LTI estimators and controllers by using extension. A LTI estimator for the system Σ can be described by:

$$\Gamma_e: \begin{array}{l} \dot{z} = Fz + Gy + Eu, \\ v = Hz + Ky, \end{array} \tag{13}$$

where $z \in \mathbb{R}^p$ is the state and $v \in \mathbb{R}^k$ is the output of the estimator Γ_e; the output v of Γ_e is an estimate of w given in (7).

A LTI controller for the system Σ can be described by:

$$\Gamma_c: \begin{array}{l} \dot{z} = Fz + Gy, \\ v = Hz + Ky, \end{array} \tag{14}$$

where $z \in \mathbb{R}^p$ is the state and $v \in \mathbb{R}^m$ is the output of the controller Γ_c; the output v of Γ_c is applied to the input of the system Σ in order to control it; i.e.: $u = v$.

A LTI estimator for the system $\tilde{\Sigma}$ can be described by:

$$\tilde{\Gamma}_e:\quad \begin{aligned}\dot{\tilde{z}} &= \tilde{F}\tilde{z} + \tilde{G}\tilde{y},\\ \tilde{v} &= \tilde{H}\tilde{z} + \tilde{K}\tilde{y},\end{aligned} \tag{15}$$

where $\tilde{z} \in \mathbb{R}^{\tilde{p}}$ is the state and $\tilde{v} \in \mathbb{R}^{k}$ is the output of the estimator $\tilde{\Gamma}_e$; the output $\tilde{v}$ is an *estimate* of $\tilde{w}$ given in (8). Similarly, a LTI controller for the system $\tilde{\Sigma}$ can be described by:

$$\begin{aligned}\dot{\tilde{z}} &= \tilde{F}\tilde{z} + \tilde{G}\tilde{y},\\ \tilde{v} &= \tilde{H}\tilde{z} + \tilde{K}\tilde{y},\end{aligned} \tag{16}$$

where $\tilde{z} \in \mathbb{R}^{p}$ is the state and $\tilde{v} \in \mathbb{R}^{\tilde{m}}$ is the output of the controller $\tilde{\Gamma}_c$; the output $\tilde{v}$ is applied to the input of the system $\tilde{\Sigma}$ for control purposes: $\tilde{u} = \tilde{v}$.

It is assumed that the dimensions of Γ_e and Γ_c do not exceed the dimensions of $\tilde{\Gamma}_e$ and $\tilde{\Gamma}_c$, respectively (i.e. $\tilde{p} \geq p$ in both cases); this assumption may be justified since Σ is a part of $\tilde{\Sigma}$, and thus should not require a larger dimensional estimator or controller (Ikeda and Šiljak, 1986). In addition to the transformations (3a)–(3e) and (9), we define the following transformation:

$$P: \mathbb{R}^{\tilde{p}} \rightarrow \mathbb{R}^{p}, \quad \operatorname{rank}(P) = p. \tag{17}$$

We now introduce contractibility for estimators and controllers:

Definition 3a. The estimator (15) is contractible to the estimator (13) if there exist transformations as in (3a), (3e), (9), and (17) such that for any initial state $x_0 \in \mathbb{R}^{n}$ of the system Σ, for any input $u(t) \in \mathbb{R}^{m}$, $0 \leq t < \infty$, of the system Σ, and for any initial state $\tilde{z}_0 \in \mathbb{R}^{\tilde{p}}$ of the estimator $\tilde{\Gamma}_e$ the choice

$$\tilde{x}_0 = Tx_0, \tag{18a}$$

$$\tilde{u}(t) = R^{\#}u(t), \quad \forall t \geq 0, \tag{18b}$$

and

$$z_0 = P\tilde{z}_0, \tag{18c}$$

implies that

$$z(t; z_0, y, u) = P\tilde{z}(t; \tilde{z}_0, \tilde{y}, \tilde{u}), \quad \forall t \geq 0, \tag{19a}$$

and

$$v(t; z, y) = Q\tilde{v}(t; \tilde{z}, \tilde{y}), \quad \forall t \geq 0. \tag{19b}$$

Definition 3b. The controller (16) is contractible to the controller (14) if there exist transformations as in (3a), (3b), and (17) such that for any initial state $x_0 \in \mathbb{R}^{n}$ of the system Σ, for any input $\tilde{u}(t) \in \mathbb{R}^{\tilde{m}}$, $0 \leq t < \infty$, of the system $\tilde{\Sigma}$, and for any initial state $\tilde{z}_0 \in \mathbb{R}^{\tilde{p}}$ of the controller $\tilde{\Gamma}_c$ the choice

$$\tilde{x}_0 = Tx_0, \tag{20a}$$

$$u(t) = R\tilde{u}(t), \quad \forall t \geq 0, \tag{20b}$$

and

$$z_0 = P\tilde{z}_0, \tag{20c}$$

implies that

$$z(t; z_0, y, u) = P\tilde{z}(t; \tilde{z}_0, \tilde{y}, \tilde{u}), \quad \forall t \geq 0, \tag{21a}$$

and

$$v(t; z, y) = R\tilde{v}(t; \tilde{z}, \tilde{y}), \quad \forall t \geq 0. \tag{21b}$$

Remark 2. Note that contractibility for an estimator is defined for an arbitrary input $u(t)$ in the original input space, and the input $\tilde{u}(t)$ for the expanded system $\tilde{\Sigma}$ is obtained by the transformation (18b). This is the natural choice, since the actual input applied to the real system Σ is $u(t)$. In this case, there is no restriction on the input $u(t)$ which is assumed to be applied to the fictitious system $\tilde{\Sigma}$. In the case of a controller, however, since the controller is designed in the expanded spaces, $\tilde{u}(t)$ must be arbitrary (otherwise, there will be a restriction on the controllers that could be designed, e.g. see Ikeda and Šiljak, 1986). Therefore, the contractibility, in this case, must be defined for an arbitrary $\tilde{u}(t)$ in the expanded input space. The outputs $v(t)$ of Γ_c and $\tilde{v}(t)$ of $\tilde{\Gamma}_c$ must be such that when the loops are closed (i.e. when $u(t) = v(t)$ and $\tilde{u}(t) = \tilde{v}(t)$) the condition (20b) is satisfied. The requirement (21b) ensures this condition.

It is important to satisfy contractibility in order to preserve the desired relations (such as stability, good performance, and good estimation) between the expanded and the original systems following the application of the appropriate estimators or controllers. First, let us introduce the following definitions.

Definition 4a. Given a system (such as Σ), we say that an estimator (such as Γ_e) achieves good estimation with respect to a given weight $W \in \mathbb{S}^{k}$ and a given scalar tolerance function $g(t)$ if $\|v(t) - w(t)\|_W \leq g(t)$, $\forall t \geq 0$, where $v(t)$ is the output of the estimator, and $w(t)$ is the output of the system which is to be estimated.

Definition 4b. Given a system (such as Σ), we say that a controller (such as Γ_c) achieves good performance with respect to a given trajectory function $f(t) \in \mathbb{R}^{n}$, $0 \leq t < \infty$, a given weight $W \in \mathbb{S}^{n}$, and a given scalar *tolerance function* $g(t)$ if $\|x(t) - f(t)\|_W \leq g(t)$, $\forall t \geq 0$, where $x(t) \in \mathbb{R}^{n}$ is the state of the controlled system.

Now we can prove the following results:

Theorem 3a. Suppose that the output (8) of the system $\tilde{\Sigma}$ includes the output (7) of the system Σ and that the estimator $\tilde{\Gamma}_e$ is contractible to the estimator Γ_e. Moreover, suppose that the estimator Γ_e is applied to the system Σ, that the estimator $\tilde{\Gamma}_e$ is applied to the system $\tilde{\Sigma}$, and that (18a)–(18c) hold. Then:

(a) If $\tilde{\Gamma}_e$ achieves asymptotic estimation then Γ_e achieves asymptotic estimation, i.e. if $\lim_{t\rightarrow\infty}(\tilde{v} - \tilde{w}) = 0$ then $\lim_{t\rightarrow\infty}(v - w) = 0$.

(b) If $\tilde{\Gamma}_e$ achieves good estimation with respect to $Q^{T}WQ$ and $g(t)$, then Γ_e achieves good estimation with respect to W and $g(t)$.

Proof. (a) Note that, by (19b) $Q\tilde{v} = v$, and by (10) $Q\tilde{w} = w$; therefore, $(v - w) = Q(\tilde{v} - \tilde{w})$, from which the desired result follows.

(b) Note that, $\|\tilde{v}(t) - \tilde{w}(t)\|_{Q^{T}WQ} \leq g(t)$ implies $\|Q(\tilde{v}(t) - \tilde{w}(t))\|_W \leq g(t)$. We have already shown that $Q(\tilde{v} - \tilde{w}) = (v - w)$; thus $\|v(t) - w(t)\|_W \leq g(t)$. □

Theorem 3b. Suppose that $\tilde{\Sigma}$ is an extension of Σ and that the controller $\tilde{\Gamma}_c$ is contractible to the controller Γ_c. Moreover, suppose that the controller Γ_c is applied to the system Σ and that the controller $\tilde{\Gamma}_c$ is applied to the system $\tilde{\Sigma}$. Then:

(a) Stability (respectively asymptotic stability) of the expanded closed-loop system (obtained by applying $\tilde{\Gamma}_c$ to $\tilde{\Sigma}$) implies the stability (respectively asymptotic stability) of the original closed-loop system (obtained by applying Γ_c to Σ);

(b) Assuming that $\tilde{x}_0 = Tx_0$ and $z_0 = P\tilde{z}_0$, if $\tilde{\Gamma}_c$ achieves good performance with respect to $\tilde{f}(t) \triangleq Tf(t)$, $\tilde{W} \triangleq (T^{\#})^{T}WT^{\#}$, and $g(t)$ then Γ_c achieves good performance with respect to $f(t)$, W, and $g(t)$.

Proof. (a) If the expanded closed-loop system is stable (asymptotically stable), then the overall state vector $[\tilde{x}(t; \tilde{x}_0, \tilde{z}_0)^{T}\tilde{z}(t; \tilde{x}_0, \tilde{z}_0)^{T}]^{T}$ of the closed-loop expanded system is bounded for all $t \geq 0$ (and $\lim_{t\rightarrow\infty}[\tilde{x}(t; \tilde{x}_0, \tilde{z}_0)^{T}\tilde{z}(t; \tilde{x}_0, \tilde{z}_0)^{T}]^{T} = 0$) for all $\tilde{x}_0 \in \mathbb{R}^{\tilde{n}}$ and for all $\tilde{z}_0 \in \mathbb{R}^{\tilde{p}}$. By (5a) and (21a), the overall state vector of the original system has the property:

$$\begin{bmatrix} x(t; x_0, z_0) \\ z(t; x_0, z_0) \end{bmatrix} = \begin{bmatrix} T^{\#} & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} \tilde{x}(t; Tx_0, P^{\#}z_0) \\ \tilde{z}(t; Tx_0, P^{\#}z_0) \end{bmatrix}, \quad \forall t \geq 0, \tag{22}$$

where $P^{\#}$ satisfies $PP^{\#} = I_p$, and thus,

$$[x(t; x_0, z_0)^{T}z(t; x_0, z_0)^{T}]^{T},$$

is also bounded for all $t \geq 0$ (and

$$\lim_{t\rightarrow\infty}[x(t; x_0, z_0)^{T}z(t; x_0, z_0)^{T}]^{T} = 0),$$

for all $x_0 \in T^{\#}\tilde{\mathbb{R}}^{\tilde{n}} = \mathbb{R}^n$ and for all $z_0 \in P\mathbb{R}^{\tilde{p}} = \mathbb{R}^p$, which implies that the original closed-loop system is also stable (asymptotically stable).

(b) If $\|\tilde{x}(t) - \tilde{f}(t)\|_{\tilde{W}} \le g(t)$ then $\|T^{\#}\tilde{x}(t) - T^{\#}\tilde{f}(t)\|_W \le g(t)$; furthermore, since $T^{\#}T = I_n$, $T^{\#}\tilde{f}(t) = f(t)$ and (5a) implies that $T^{\#}\tilde{x}(t) = x(t)$; thus $\|x(t) - f(t)\|_W \le g(t)$. □

Remark 3. Recall that the estimator or the controller is to be designed in the expanded spaces and then contracted for implementation. Thus, in the case of an estimator, if good estimation with respect to W and $g(t)$ and/or asymptotic estimation is desired for the original system under inputs $u(t) \in \mathscr{U} \subset \mathbb{R}^m$ and initial conditions $x_0 \in \mathscr{X} \subset \mathbb{R}^n$ and $z_0 \in \mathscr{Z} \subset \mathbb{R}^p$, then the estimator $\tilde{\Gamma}_e$ must be designed to achieve good estimation with respect to Q^TWQ and $g(t)$ and/or asymptotic estimation for the expanded system under inputs $\tilde{u}(t) \in \tilde{\mathscr{U}} \supset R^{\#}\mathscr{U}$ and initial conditions $\tilde{x}_0 \in \tilde{\mathscr{X}} \supset T\mathscr{X}$ and $\tilde{z}_0 \in \tilde{\mathscr{Z}} \supset P^{\#}\mathscr{Z}$. Similarly, in the case of a controller, if closed-loop (asymptotic) stability and good performance with respect to $f(t)$, W, and $g(t)$ are desired for the original system under initial conditions $x_0 \in \mathscr{X} \subset \mathbb{R}^n$ and $z_0 \in \mathscr{Z} \subset \mathbb{R}^p$, then the controller $\tilde{\Gamma}_c$ must be designed to achieve closed-loop (asymptotic) stability and good performance with respect to $Tf(t)$, $(T^{\#})^TWT^{\#}$, and $g(t)$ for the expanded system under initial conditions $\tilde{x}_0 \in \tilde{\mathscr{X}} \supset T\mathscr{X}$ and $\tilde{z}_0 \in \tilde{\mathscr{Z}} \supset P^{\#}\mathscr{Z}$.

The necessary and sufficient conditions for the contractibility of estimators and controllers are given by the following theorems.

Theorem 4a. The estimator (15) is contractible to the estimator (13) if and only if there exist transformations as in (3a), (3e), (9), and (17) such that

$$FP = P\tilde{F}, \tag{23a}$$

$$GCA^j = P\tilde{G}\tilde{C}\tilde{A}^jTM \quad \forall j \in \{0, 1, 2, \ldots\}, \tag{23b}$$

$$GCA^jB = P\tilde{G}\tilde{C}\tilde{A}^j\tilde{B}R^{\#}, \quad \forall j \in \{0, 1, 2, \ldots \tag{23c}$$

$$E = P\tilde{E}R^{\#}, \tag{23d}$$

$$HP = Q\tilde{H}, \tag{23e}$$

$$KCA^j = Q\tilde{K}\tilde{C}\tilde{A}^jT, \quad \forall j \in \{0, 1, 2, \ldots\}, \tag{23f}$$

and

$$KCA^jB = Q\tilde{K}\tilde{C}\tilde{A}^j\tilde{B}R^{\#}, \quad \forall j \in \{0, 1, 2, \ldots \tag{23g}$$

Proof. The proof follows similar lines to the proof of Theorem 1. □

Theorem 4b. The controller (16) is contractible to the controller (14) if and only if there exist transformations as in (3a), (3b), and (17) such that

$$FP = P\tilde{F}, \tag{24a}$$

$$GCA^j = P\tilde{G}\tilde{C}\tilde{A}^jT, \quad \forall j \in \{0, 1, 2, \ldots\} \tag{24b}$$

$$GCA^jBR = P\tilde{G}\tilde{C}\tilde{A}^j\tilde{B}, \quad \forall j \in \{0, 1, 2, \ldots\} \tag{24c}$$

$$HP = R\tilde{H}, \tag{24d}$$

$$KCA^j = R\tilde{K}\tilde{C}\tilde{A}^jT, \quad \forall j \in \{0, 1, 2, \ldots\} \tag{24e}$$

and

$$KCA^jBR = R\tilde{K}\tilde{C}\tilde{A}^j\tilde{B}, \quad \forall j \in \{0, 1, 2, \ldots\}. \tag{24f}$$

Proof. The proof follows similar lines to the proof of Theorem 1. □

The above conditions reduce to simpler ones if the expanded system is an extension of the original system.

Corollary 2a. Given that the system $\tilde{\Sigma}$ is an extension of the system Σ, the estimator (15) for the system $\tilde{\Sigma}$ is contractible to the estimator (13) for the system Σ if there exist transformations as in (9) and (17) such that

$$FP = P\tilde{F}, \tag{25a}$$

$$GC = P\tilde{G}SC, \tag{25b}$$

$$E = P\tilde{E}R^{\#}, \tag{25c}$$

$$HP = Q\tilde{H}, \tag{25d}$$

and

$$KC = Q\tilde{K}SC, \tag{25e}$$

where S is the transformation satisfying (6c) and $R^{\#}$ is a matrix satisfying $RR^{\#} = I_m$, where R is the transformation satisfying (6b).

Proof. Given (6a)–(6c), it can be shown that (23b) and (23c) both reduce to (25b), and (23f) and (23g) both reduce to (25e). □

Corollary 2b. Given that the system $\tilde{\Sigma}$ is an extension of the system Σ, the controller (16) for the system $\tilde{\Sigma}$ is contractible to the controller (14) for the system Σ if there exists a transformation as in (17) such that

$$FP = P\tilde{F}, \tag{26a}$$

$$GC = P\tilde{G}SC, \tag{26b}$$

$$HP = R\tilde{H}, \tag{26c}$$

and

$$KC = R\tilde{K}SC, \tag{26d}$$

where S and R are the transformations satisfying (6c) and (6b), respectively:

Proof. Given (6a)–(6c), it can be shown that (24b) and (24c) both reduce to (26b), and (24e) and (24f) both reduce to (26d). □

Since the estimator/controller is to be designed in the expanded spaces and then contracted for implementation, it is important that any estimator/controller designed in the expanded spaces be contractible. In fact, if $\tilde{\Sigma}$ is an extension of Σ then such a property holds.

Corollary 3a. If the system $\tilde{\Sigma}$ is an extension of the system Σ then any estimator of the form (15) for the system $\tilde{\Sigma}$ is contractible to an estimator of the form (13) for the system Σ with:

$$F = \tilde{F}, \quad G = \tilde{G}S, \quad E = \tilde{E}R^{\#}, \quad H = Q\tilde{H}, \quad K = Q\tilde{K}S. \tag{27}$$

Proof. With $P = I_{\tilde{p}}$, F, G, E, H, and K defined above satisfy (25a)–(25e). □

Corollary 3b. If the system $\tilde{\Sigma}$ is an extension of the system Σ, then any controller of the form (16) for the system $\tilde{\Sigma}$ is contractible to a controller of the form (14) for the system Σ with:

$$F = \tilde{F}, \quad G = \tilde{G}S, \quad H = R\tilde{H}, \quad K = R\tilde{K}S. \tag{28}$$

Proof With $P = I_p$, F, G, H, and K defined above satisfy (26a)–(26d). □

Remark 4. Hodžić and Šiljak (1986) have suggested the use of aggregation for the design of estimators for large scale systems. They showed that, for the case of state expansion only (i.e. when the original and expanded input and output spaces are identical), if the original system (1) is an aggregation of the expanded system (2) then any estimator of the form

$$\dot{\hat{\tilde{x}}} = \tilde{A}\hat{\tilde{x}} + \tilde{B}\tilde{u} + \tilde{K}[\tilde{y} - \tilde{C}\hat{\tilde{x}}], \tag{29}$$

for the expanded system is contractible to an estimator of the form

$$\dot{\hat{x}} = A\hat{x} + Bu + K[y - C\hat{x}], \tag{30}$$

for the original system. Here $\hat{\tilde{x}} \in \mathbb{R}^{\tilde{n}}$ and $\hat{x} \in \mathbb{R}^n$ are the estimates of the state vectors of the expanded and the original systems, respectively. Corollary 3a, on the other

hand, shows that if extension is used then any estimator of the form (15) (which is a more general form than (29)) can be contracted for implementation. The use of the more general form ((15) over the form (29)) may be useful, since it allows the design of decentralized estimators (note that (29) is not a decentralized estimator unless the system matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ can be obtained as block diagonal matrices, which can happen only in trivial cases). As will be illustrated in the next section, decentralized estimators for the expanded system that are designed using extension result in estimators for the original system which have a special overlapping-decentralized structure. Such a structure may be useful in the implementation of the estimators (the estimator (30) lacks this structure unless the system matrices A, B, and C have a special structure).

4. *Overlapping decompositions*

In this section decentralized estimator and controller design with overlapping decompositions is discussed within the framework of extension. For notational simplicity, only systems with two decentralized estimation/control agents are considered. The extension of the results to systems with more agents is straightforward.

Consider the system Σ given in (1). Suppose that the state, the input, and the output are partitioned as:

$$x = (x_1^T, x_2^T, x_3^T)^T, \quad x_i \in \mathbb{R}^{n_i} \qquad i = 1, 2, 3, \tag{31a}$$

$$u = (u_1^T, u_2^T, u_3^T)^T, \quad u_i \in \mathbb{R}^{m_i} \qquad i = 1, 2, 3, \tag{31b}$$

and

$$y = (y_1^T, y_2^T, y_3^T)^T, \quad y_i \in \mathbb{R}^{l_i}, \quad i = 1, 2, 3. \tag{31c}$$

Here it is assumed that x_2, u_2, and y_2 correspond to the overlapping parts of the state, input, and output spaces, respectively. Then, by choosing appropriate transformations and complementary matrices, an extension $\tilde{\Sigma}$, as described in (2), of Σ can be obtained (for details see İftar, 1990). If an estimator is to be designed, we also partition the variable to be estimated as follows:

$$w = (w_1^T, w_2^T, w_3^T)^T, \quad w_i \in \mathbb{R}^{k_i}, \quad i = 1, 2, 3. \tag{32}$$

The matrix D in (7) is also partitioned compatibly:

$$D = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix}$$

The corresponding variable to be estimated for the expanded system is given by (8), where, for an appropriate transformation Q,

$$\tilde{D} = \begin{bmatrix} D_{11} & D_{12} & 0 & D_{13} \\ D_{21} & \frac{1}{2}D_{22} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \frac{1}{2}D_{22} & D_{23} \\ D_{31} & 0 & D_{32} & D_{33} \end{bmatrix}$$

Suppose that local estimators described by:

$$\tilde{\Gamma}_e^i: \begin{array}{l} \dot{\tilde{z}}_i = \tilde{F}_i\tilde{z}_i + \tilde{G}_i\tilde{y}_i + \tilde{E}_i\tilde{u}_i, \\ \tilde{v}_i = \tilde{H}_i\tilde{z}_i + \tilde{K}_i\tilde{y}_i \end{array} \qquad i = 1, 2. \tag{33}$$

are designed such that the overall expanded estimator $\tilde{\Gamma}_e$ described by (15), with

$$\tilde{F} = \text{blockdiag}\,\{\tilde{F}_1, \tilde{F}_1\}, \qquad \tilde{G} = \text{blockdiag}\,\{\tilde{G}_1, \tilde{G}_2\}, \tag{34a}$$

$$\tilde{H} = \text{blockdiag}\,\{\tilde{H}_1, \tilde{H}_2\}, \qquad \tilde{E} = \text{blockdiag}\,\{\tilde{E}_1, \tilde{E}_2\}, \tag{34b}$$

and

$$\tilde{K} = \text{blockdiag}\,\{\tilde{K}_1, \tilde{K}_2\}, \tag{34c}$$

satisfies the design requirements (such as asymptotic estimation or good estimation) when applied to the expanded system $\tilde{\Sigma}$. Let us partition the matrices of the local estimators as follows:

$$\tilde{G}_i = [G_1^i \quad G_2^i], \quad \tilde{E}_i = [E_1^i \quad E_2^i], \tag{35a}$$

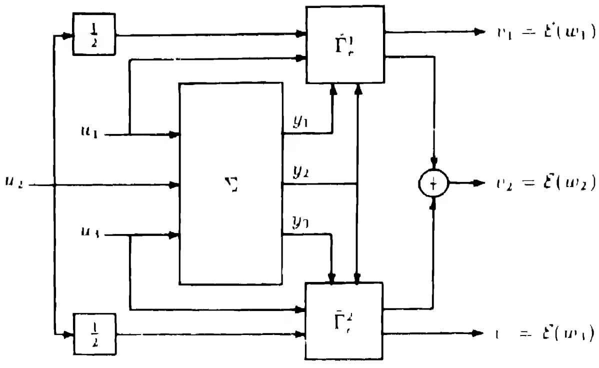


FIG. 1. Implementation of overlapping decentralized estimators.

$$\tilde{H}_i = \begin{bmatrix} H_1^i \\ H_2^i \end{bmatrix}, \quad \tilde{K}_i = \begin{bmatrix} K_{11}^i & K_{12}^i \\ K_{21}^i & K_{22}^i \end{bmatrix}, \quad i = 1, 2, \tag{35b}$$

where $G_1^1 \in \mathbb{R}^{\bar{p}_1 \times l_1}$, $G_2^2 \in \mathbb{R}^{\bar{p}_2 \times l_3}$, $E_1^1 \in \mathbb{R}^{\bar{p}_1 \times m_2}$, $E_2^2 \in \mathbb{R}^{\bar{p}_2 \times m_3}$, $H_1^1 \in \mathbb{R}^{k_1 \times \bar{p}_1}$, $H_2^2 \in R^{k_3 \times \bar{p}_2}$, $K_{11}^1 \in \mathbb{R}^{k_1 \times l_1}$, $K_{22}^2 \in \mathbb{R}^{k_3 \times l_3}$, and $\bar{p}_i$ is the dimension of $\tilde{z}_i$ ($i = 1, 2$). The contraction of the above described overall expanded estimator to the original spaces is given by (13), where

$$F = \tilde{F}, \quad G = \begin{bmatrix} G_1^1 & G_2^1 & 0 \\ 0 & G_1^2 & G_2^2 \end{bmatrix},$$

$$E = \begin{bmatrix} E_1^1 & \frac{1}{2}E_2^1 & 0 \\ 0 & \frac{1}{2}E_1^2 & E_2^2 \end{bmatrix}, \tag{36a}$$

$$H = \begin{matrix} H_1^1 & \\ H_2^1 & H_1^2 \\ 0 & H_2^2 \end{matrix} \qquad K = \begin{matrix} K_{21}^1 & K_{22}^1 + K_{11}^2 & K_{12}^2 \\ 0 & K_{21}^2 & \end{matrix} \tag{36b}$$

The implementation of these overlapping decentralized estimators (which form the contracted estimator Γ_e) is illustrated in Fig. 1. Assuming that the decentralized estimator $\tilde{\Gamma}_e$ achieves asymptotic estimation and/or good estimation when applied to the expanded system $\tilde{\Sigma}$, by Theorem 3a, the above described contracted estimator Γ_e achieves asymptotic estimation and/or good estimation when applied to the original system Σ.

Next, suppose that local controllers described by:

$$\tilde{\Gamma}_c^i: \begin{array}{l} \dot{\tilde{z}}_i = \tilde{F}_i\tilde{z}_i + \tilde{G}_i\tilde{y}_i \\ \tilde{v}_i = \tilde{H}_i\tilde{z}_i + \tilde{K}_i\tilde{y}_i \end{array}, \quad i = 1, 2, \tag{37}$$

are designed such that the overall expanded controller $\tilde{\Gamma}_c$ described by (16), with

$$\tilde{F} = \text{blockdiag}\,\{\tilde{F}_1, \tilde{F}_2\}, \qquad \tilde{G} = \text{blockdiag}\,\{\tilde{G}_1, \tilde{G}_2\}, \tag{38a}$$

$$\tilde{H} = \text{blockdiag}\,\{\tilde{H}_1, \tilde{H}_2\}, \qquad \tilde{K} = \text{blockdiag}\,\{\tilde{K}_1, \tilde{K}_2\}, \tag{38b}$$

satisfies the design requirements (such as stability and/or good performance) when applied to the expanded system $\tilde{\Sigma}$. Let the matrices of the local controllers be partitioned in a similar way to the partitioning of the local estimator matrices given in (35a)–(35b). The contraction of this overall expanded controller to the original spaces is then given by (14), where the matrices F, G, H, and K are obtained from the matrices of $\tilde{\Gamma}_c^1$ and $\tilde{\Gamma}_c^2$ as described in (36a)–(36b). The implementation of this contracted controller, which consists of overlapping decentralized controllers, is illustrated in Fig. 2. Assuming that the designed decentralized controller $\tilde{\Gamma}_c$ achieves stability and/or good performance when applied to the expanded system $\tilde{\Sigma}$, by Theorem 3b, the above described contracted controller Γ_c achieves stability and/or good performance when applied to the original system Σ.

Remark 5. Comparing the contracted controllers obtained in this section and those obtained by Ikeda and Šiljak (1986), it is seen that they both have the same structure. However, there is an important difference in the intermediate design stages. Since unrestriction is used in Ikeda and Šiljak (1986),

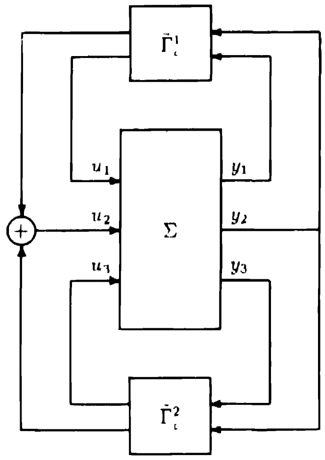


FIG. 2. Implementation of overlapping decentralized controllers.

the controller designed for the expanded system may not be contractible for implementation on the original system. It is, therefore, required to modify the designed controller (as shown in Ikeda and Šiljak, 1986) before contraction. This modified controller, however, need not satisfy the design requirements (e.g. stability or good performance) even if the unmodified controller does. Therefore, either the expanded closed-loop system with the modified controller or the original closed-loop system with the contracted controller must be tested for these requirements; and the design must be repeated if these requirements are not satisfied. On the other hand, the controllers designed using the present approach can be contracted directly without the need of any modifications, testing, or redesign, since the contractibility of these controllers is guaranteed by Corollary 3b.

5. *Conclusions*

The extension principle has been generalized to the case where the output space is also expanded besides the state and the input spaces. This principle is conceptually close to the unrestriction defined by Ikeda and Šiljak (1986). However, the unrestriction was defined for an arbitrary input in the original input space, and the input in the expanded space was obtained by a transformation. This fact brings in the restriction that a particular estimator or controller designed for the expanded system by using unrestriction may not be contractible to the original spaces. It has been shown, however, that with the approach undertaken here any estimator or controller designed in the expanded spaces is always contractible to the original spaces.

Furthermore, it has been shown that, if an estimator designed for the expanded system achieves good estimation and/or asymptotic estimation then the contracted estimator also achieves good estimation and/or asymptotic estimation for the original system. Similarly, if a controller designed for the expanded system achieves stability and/or good performance then the contracted controller also achieves stability and/or good performance for the original system.

Finally, decentralized estimator and controller design with overlapping decompositions has been discussed within the framework of extension. Although systems with only two decentralized agents have been presented for notational simplicity, the extension of the results to systems with more than two agents is straightforward.

Acknowledgement—The author wishes to thank the referees for their constructive comments.

References

Hodžić, M. and D. D. Šiljak (1986). Decentralized estimation and control with overlapping information sets. *IEEE Trans. Aut. Control*, **AC-31**, 83–86.

İftar, A. (1990). Decentralized estimation and control with overlapping input, state, and output decomposition. *Preprints of IFAC World Congress*, Vol. 2, pp. 13–18, Tallinn, Estonia.

İftar, A. and Ü. Özgüner (1987). Local LQG/LTR controller design for decentralized systems. *IEEE Trans. Aut. Control*, **AC-32**, 926–930.

İftar, A. and Ü. Özgüner (1990). Contractible controller design and optimal control with state and input inclusion. *Automatica*, **26**, 593–597.

Ikeda, M. and D. D. Šiljak (1980). Overlapping decompositions, expansions, and contractions of dynamic systems. *Large Scale Systems*, **1**, 29–38.

Ikeda, M. and D. D. Šiljak (1986). Overlapping decentralized control with input, state, and output inclusion. *Control Theory and Advanced Technology*, **2**, 155–172.

Ikeda, M., D. D.Šiljak and D. E. White (1984). An inclusion principle for dynamic systems. *IEEE Trans. Aut. Control*, **AC-29**, 244–249.

Ohta, Y., D. D. Šiljak and T. Matsumoto (1986). Decentralized control using quasi-block diagonal dominance of transfer function matrices. *IEEE Trans. Aut. Control*, **AC-31**, 420–430.

Özgüner, Ü., F. Khorrami and A. İftar (1988). Two controller design approaches for decentralized systems. *Proc. of the AIAA Guidance, Navigation, and Control Conf.*, pp. 337–344, Minneapolis, MN.

Appendix

Proof of Theorem 1. The responses of the two systems (1) and (2) are given by

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)\,d\tau, \tag{39}$$

and

$$\tilde{x}(t) = e^{\tilde{A}t}\tilde{x}_0 + \int_0^t e^{\tilde{A}(t-\tau)}\tilde{B}\tilde{u}(\tau)\,d\tau, \tag{40}$$

respectively. By substituting (4a) and (4b) for $\tilde{x}_0$ and $u(\tau)$, respectively, and using the power series expansion for the matrix exponentials, it can be shown that (5a) holds for any x_0 and any $\tilde{u}(t)$ if and only if

$$\tilde{A}^iT = TA^i, \tag{41a}$$

and

$$\tilde{A}^i\tilde{B} = TA^iBR, \tag{41b}$$

for all $i \in \{0, 1, 2, \ldots\}$. Similarly, (5b) holds for any x_0 and any $\tilde{u}(t)$ if and only if

$$\tilde{C}\tilde{A}^iT = SCA^i, \tag{42a}$$

and

$$\tilde{C}\tilde{A}^i\tilde{B} = SCA^iBR, \tag{42b}$$

for all $i \in \{0, 1, 2, \ldots\}$. Now it is straightforward to show the only if part, since (6a) can be obtained from (41a) with $i = 1$, (6b) can be obtained from (41b) with $i = 0$, and (6c) can be obtained from (42a) with $i = 0$. To show the if part, note that for $i = 0$ (41a) holds trivially; for $i = 1$ (41a) holds if and only if (6a) holds, for $i = 0$ (41b) holds if and only if (6b) holds; and if (6a) holds, then (41a) holds for $i \geq 2$ and (41b) holds for $i \geq 1$. Similarly, for $i = 0$, (42a) holds if and only if (6c) holds, and given that (6b) holds, (42b) holds if (6c) holds, and for $i \geq 1$, (42a) and (42b) hold if (6a)–(6c) hold. Hence, the result follows.

Automatica, Vol. 29, No. 2, pp. 517–522, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Brief Paper

Spectral Factorization of Linear Periodic Systems with Application to the Optimal Prediction of Periodic ARMA Models*

SERGIO BITTANTI† and GIUSEPPE DE NICOLAO‡

Key Words—Discrete-time systems; time-varying systems; spectral factorization; zeros; prediction theory; Kalman filters; ARMA models; periodic systems; cyclostationary processes.

Abstract—A cyclostationary process is a stochastic process whose statistical parameters, such as mean and autocorrelation, exhibit suitable periodicity. In this paper, we consider the cyclospectral factorization problem which consists of finding a Markovian (state-space) realization of a given cyclostationary process. It is shown that a significant class of periodic state-space representations is in a one-to-one correspondence with the periodic solutions of a difference periodic Riccati equation. This result is applied to the solution of the prediction problem for ARMA models with periodically varying coefficients. If the periodic ARMA model is minimum-phase, the optimal predictor is given a simple input–output expression that generalizes the well-known one for time-invariant ARMAs. Otherwise, the computation of the optimal predictor calls for the solution of a cyclospectral factorization problem.

1. *Introduction*

A CYCLOSTATIONARY (CS) process is characterized by the fact that its statistical parameters (mean, variance, autocorrelation) are periodically time-varying (Gardner and Franks, 1975). Processes of this type naturally arise whenever some form of periodicity is involved and their applications range from the modeling and prediction of seasonal phenomena, see e.g. Vecchia (1985), to the synthesis of communication systems (Gardner, 1991). In trying to develop a system theory approach to the study of such processes, they have been characterized as the output of a linear system (fed by white noise) whose impulse response exhibits some suitable periodicity (Gardner, 1975) or alternatively, as the output of linear input–output models with periodic coefficients (periodic ARMA models) (Jones and Brelsford, 1967; Pagano, 1978). So, apart from basic issues dealt with in Bittanti (1986), little is known about the Markovian (state-space) representation of CS processes. In particular, it would be highly desirable to characterize the set of Markovian representations of the same CS process as well as investigate the existence of innovations representations leading, in analogy with the stationary case, to stable whitening filters.

The main objective of this paper is just to lay the foundations of a state-space theory of discrete-time CS processes. In particular we will examine the following problem: given a CS process which has a finite-dimensional linear periodic realization, characterize a whole set of alternative periodic state-space representations of the process. The main result of the paper is that there exists a bijective correspondence between the innovations-like representations and the periodic solutions of a certain Difference Periodic Riccati Equation (DPRE). Furthermore, by making reference to the notion of zeros of a periodic system, it is shown that there exists a unique minimum-phase (canonical) representation corresponding to the (unique) stabilizing periodic solution of the DPRE.

Besides being of independent interest, such a "cyclospectral" factorization theory is shown (Section 5) to play a key role in the prediction problem for (univariate) ARMA models with periodic coefficients, i.e.

$$y(t) = \sum_{i=1}^{n} a_i(t-i)y(t-i) + \sum_{i=1}^{n} c_i(t-i)e(t-i) + e(t), \quad (1)$$

where $a_i(\cdot)$ and $c_i(\cdot)$ are periodically varying parameters:

$$a_i(t) = a_i(t+T), \quad c_i(t) = (t+T),$$

and $e(\cdot)$ is a White Gaussian Noise (WGN) with nonsingular periodic variance:

$$e(\cdot) \sim \mathrm{WGN}(0, R(\cdot)), \quad R(t) = R(t+T), \quad R(t) \neq 0, \quad \forall t.$$

In the paper, we will provide a complete solution to the previous problem for both minimum- and nonminimum-phase PARMA models.

There is a strict parallel arrangement between our theory and spectral factorization theory for stationary stochastic processes. Now, it is well known that discrete-time periodic systems admit an alternative representation in terms of a time invariant system with augmented input and output vectors (lifted representation) see Section 2. This kind of isomorphism between periodic and invariant systems somewhat justifies the striking analogies between spectral and cyclospectral factorization theory, and one may also expect that, by the use of the lifted representation, the discrete-time periodic results can be derived as (more or less) straightforward corollaries of the time-invariant ones. However, this is not entirely true. As a matter of fact, the analysis carried out in Bittanti and De Nicolao (1990) mainly relied on the correspondence between a periodic system and its time-invariant lifted representation. This line of reasoning allowed us to show that any positive semidefinite solution of

* Received 4 October 1990; revised 3 March 1992; received in final form 8 June 1992. A preliminary version of this paper was presented at the 11th IFAC World Congress on Automatic Control at the Service of Mankind which was held in Tallinn, Estonia during August 1992. The Published Proceedings of this IFAC Meeting may be ordered from: Pergamon Press Ltd, Headington Hill Hall, Oxford, OX3 0BW, U.K. This paper was recommended for publication in revised form by Associate Editor T. Başar under the direction of Editor H. Kwakernaak.

† Dipartimento di Elettronica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy. Fax: +39.2.23993587, email: BITTANTI@IPMEL1.POLIMI.IT.

‡ Centro Teoria dei Sistemi-CNR, c/o Dipartimento di Elettronica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy. Fax: +39.2.23993587, email: DENICOLAO@IPMEL1.POLIMI.IT.

a suitable DPRE gives rise to a representation of the given CS process, but failed to extend the same result to negative semidefinite and nondefinite periodic solutions. Consequently, no bijective correspondence could be established between periodic solutions and innovation-like representations.

Another possible approach to the study of the cyclospectral factorization problem could consist of specializing to the periodic case known results in the spectral factorization of nonstationary processes, see e.g. Anderson *et al.* (1969), Kailath (1969), Anderson and Moylan (1974) and Anderson and Moore (1979). In these papers, nonstationary spectral factors are associated with the solutions of a time-varying Riccati equation. The natural way to select the periodic spectral representations out of the infinite time-varying ones is to eliminate the transient effect due to the particular choice of the initial condition of the Riccati equation. However, according to Anderson *et al.* (1969) the factorizations obtained by moving the initialization time of the Riccati equation towards the infinitely remote past, turn out to be independent of the particular initial condition. Hence, such an approach would provide only one periodic spectral factor.

In the present paper, most of the analysis is carried out by exploiting tools belonging to the theory of linear periodic systems, with the time-invariant lifted representation playing a marginal role. Such a 'periodic approach' performs better than lifted representations or the 'periodicity' of time-varying results, in that it succeeds in establishing a bijective correspondence between innovations-like representations and periodic solutions of the DPRE. The main results of the paper are the periodic extension of classic time-invariant results, such as in Lancaster *et al.* (1986), to mention a recent discrete-time reference. However, it is worth mentioning that the translation of the proofs of Lancaster *et al.* (1986) to the periodic case is not trivial. In fact, when dealing with periodically time-varying systems one cannot resort to zeta transforms, and alternative rationales in the time domain have to be worked out. Among other things, our analysis calls for a new property of the periodic solutions of the DPRE (Theorem 2).

2. *Preliminaries*

2.1. *Periodic systems and cyclostationary processes.* Consider the stochastic periodic system

$$x(t+1) = F(t)x(t) + G(t)v(t), \tag{2a}$$

$$y(t) = H(t)x(t) + w(t). \tag{2b}$$

Matrices $F(\cdot): Z \to R^{n \times n}$, $G(\cdot): Z \to R^{n \times m}$, and $H(\cdot). Z \to R^{p \times n}$, are periodic of period T; $v(\cdot)$ and $w(\cdot)$ are white noises with Gaussian distribution, zero expected value and covariance matrix;

$$\mathrm{Var}\,[v(t)'w(t)']' = \begin{bmatrix} Q(t) & S(t) \\ S(t)' & R(t) \end{bmatrix}, \quad \forall t,$$

where $Q(t) = Q(t+T)$, $S(t) = S(t+T)$, and $R(t) = R(t+T)$ are periodic matrices of appropriate dimensions, and $R(t) > 0$, $\forall t$.

Denoting by $\Psi_F(t, \tau)$ the transition matrix relative to $F(\cdot)$, let

$$\Phi_F(t) = \Psi_F(t+T, t),$$

be the monodromy matrix at t In general $\Phi_F(t)$ depends on t; however its eigenvalues, which are called characteristic multipliers turn out to be time-independent, see e.g. Bittanti (1986). The characteristic multipliers are the periodic equivalents of the poles of a time-invariant system. Indeed, system (2) is (asymptotically) stable iff all its multipliers belong to the open unit disk If this stability condition is met with, then any solution $y(\cdot)$ asymptotically converges to a zero-mean cyclostationary process (Gardner and Franks, 1975) of period T, i.e. a process $\xi(\cdot)$ such that $\gamma(\tau, t) = \gamma(\tau+T, t+T)$, where $\gamma(t, t) = E[\xi(\tau)\xi(t)']$.

In the sequel, the following observability criterion, that straightforwardly follows from Lemma 1 in Bittanti *et al.* (1986), will prove useful.

Observability criterion

System (2) is completely observable iff rank $[V(t)] = n$, $\forall t$, where

$$V(t) = [H(t)' \quad \Psi_F(t+1, t)'H(t+1)' \quad \cdots \quad \Psi_F(t+nT-1, t)'H(t+nT-1)']'. \quad \blacksquare$$

As for stabilizability and detectability, several equivalent criteria have been worked out. In particular, one of these criteria states that system (2) is stabilizable (detectable) iff its uncontrollable (unreconstructable) part is asymptotically stable. A modal notion of stabilizability can also be introduced by making reference to the stability of the so-called unreachable characteristic multipliers. Precisely, a characteristic multiplier λ is said to be $(F(\cdot), G(\cdot))$-unreachable at time τ (Bittanti and Bolzern, 1985) if there exists $\eta \neq 0$, such that

$$\Phi_F(\tau)'\eta = \lambda\eta, \quad G(j-1)'\Psi_F(\tau, j)'\eta = 0, \quad \forall j \in [\tau - T + 1, \tau].$$

For more discussion on the basics of periodic systems see Bittanti (1986).

2.2. *Zeros of periodic systems.* A consistent definition of zeros of discrete-time periodic systems has been first given in Bolzern *et al.* (1986). To be precise, consider the system

$$x(t+1) = A(t)x(t) + B(t)u(t), \tag{3a}$$

$$y(t) = C(t)x(t) + D(t)u(t), \tag{3b}$$

with $A(\cdot)$, $B(\cdot)$, $C(\cdot)$, $D(\cdot)$ periodic matrices of period T Letting

$$A_\tau = \Phi_A(\tau), \qquad A_\tau \in R^{n \times n}$$

$$B_\tau = [\Psi_A(\tau+T, \tau+1)B(\tau) \quad \Psi_A(\tau+T, \tau+2)B(\tau+1) \quad \cdots \quad B(\tau+T-1)], \quad B_\tau \in R^{n \times mT}$$

$$C_\tau = [C(\tau)' \quad \Psi_A(\tau+1, \tau)'C(\tau+1)' \quad \cdots \Psi_A(\tau+T-1, \tau)'C(\tau+T-1)']', \quad C_\tau \in R^{pT \times n}$$

$$D_\tau = \{(D_\tau)_{ij}\}, \quad i, j = 1, 2, \ldots, T, \qquad D_\tau \in R^{pT \times mT}$$

$$(D_\tau)_{ij} = \begin{cases} 0 & i < j \\ D(\tau+i-1) & i = j, \\ C(\tau+i-1)\Psi_A(\tau+i-1, \tau+j)B(\tau+j-1) & i > j \end{cases}$$

$$u_\tau(k) = [u(\tau+kT)' \quad u(\tau+kT+1)' \quad \cdots \quad u(\tau+(k+1)T-1)']'$$

consider the time-invariant system

$$x_\tau(k+1) = A_\tau x_\tau(k) + B_\tau u_\tau(k), \tag{4a}$$

$$y_\tau(k) = C_\tau x_\tau(k) + D_\tau u_\tau(k). \tag{4b}$$

Definition 1 (Bolzern *et al.*, 1986). z is a zero at time τ of system (3), if it is a zero of the time-invariant system (4), i.e. if $\det N(z) = 0$, where

$$N(z) = \begin{bmatrix} zI - A_\tau & -B_\tau \\ C_\tau & D_\tau \end{bmatrix}.$$

As discussed in Bolzern *et al.* (1986) and Grasselli and Longhi (1988), the non-zero zeros together with their multiplicities are in fact independent of τ. In Bolzern *et al.* (1986) it has also been proven that one can associate to any zero of the system a suitable initial state and input function that result in the null output. In other words, the transmission-blocking property that characterizes the zeros of linear time-invariant systems holds in the periodic case too. In Grasselli and Longhi (1988), it has been shown that zeros of periodic systems are invariant with respect to state feedback. The invariance with respect to output feedback, as stated in the following lemma, also holds (Bittanti and De Nicolao, 1992a).

Lemma 1. Consider the following periodic system generated from system (3) by means of a periodic output feedback

$$\bar{x}(t+1) = A(t)\bar{x}(t) + E(t)\bar{y}(t) + B(t)u(t), \tag{5a}$$

$$\bar{y}(t) = C(t)\bar{x}(t) + D(t)u(t), \tag{5b}$$

where $E(\cdot)$ is an arbitrary periodic matrix of suitable dimension. Then, the zeros at time τ of (3) coincide with the zeros at τ of (5).

3. *The DPRE and its periodic solutions*

Associated with system (2) is the discrete-time periodic Riccati equation

$$\begin{aligned} P(t+1) = {} & F(t)P(t)F(t)' + G(t)Q(t)G(t)' \\ & - [G(t)S(t) + F(t)P(t)H(t)'][H(t)P(t)H(t)' \\ & + R(t)]^{-1}[G(t)S(t) + F(t)P(t)H(t)']', \end{aligned} \tag{6a}$$

or, equivalently,

$$\begin{aligned} P(t+1) = {} & A(t)P(t)A(t)' + B(t)B(t)' \\ & - A(t)P(t)H(t)'[H(t)P(t)H(t)' \\ & + R(t)]^{-1}H(t)P(t)A(t)', \end{aligned} \tag{6b}$$

where

$$A(t) = F(t) - G(t)S(t)R(t)^{-1}H(t), \tag{7a}$$

$$\begin{aligned} B(t)B(t)' = {} & G(t)Q(t)G(t)' - \\ & - G(t)S(t)R(t)^{-1}S(t)'G(t)'. \end{aligned} \tag{7b}$$

Several results on the periodic solutions of such an equation are available, see Bittanti *et al* (1988a, 1990, 1991), de Souza (1989a, 1991). In the sequel, particular attention will be paid to the periodic solutions defined below, where the symbol $\hat{A}(t)$ denotes the closed-loop matrix, which can be given two equivalent expressions as follows

$$\hat{A}(t) = A(t) - \Gamma(t)H(t) = F(t) - K(t)H(t), \tag{8a}$$

$$\begin{aligned} K(t) = {} & [G(t)S(t) + F(t)P(t)H(t)'] \\ & \times [H(t)P(t)H(t)' + R(t)]^{-1}, \end{aligned} \tag{8b}$$

$$\Gamma(t) = A(t)P(t)H(t)'[H(t)P(t)H(t)' + R(t)]^{-1} \tag{8c}$$

Definition 2. A symmetric periodic solution $\hat{P}(\cdot)$ of the DPRE is said to be

maximal if $P(t) \le \hat{P}(\cdot)$, $\forall t$, for all symmetric periodic solutions $P(\cdot)$ of the DPRE;

strong if the characteristic multipliers of $\hat{A}(\cdot)$ belong to the closed unit disk;

stabilizing if the characteristic multipliers of $\hat{A}(\cdot)$ belong to the open unit disk. ■

As for the Symmetric Periodic Positive Semidefinite (SPPS) solutions, the following results can be found in the literature (Bittanti *et al.*, 1988a, 1990; de Souza, 1989b, 1991).

Theorem 1.

(1.1) If $(A(\cdot), H(\cdot))$ is detectable, then the DPRE admits a maximal symmetric periodic solution $P^+(\cdot)$, which is in fact positive semidefinite and strong.

(1.2) If $(A(\cdot)B(\cdot))$ is stabilizable and $(A(\cdot), H(\cdot))$ detectable, then the DPRE admits a unique SPPS solution, which is in fact stabilizing.

(1.3) The DPRE admits a stabilizing solution iff $(A(\cdot), H(\cdot))$ is detectable and there is no unit-modulus $(A(\cdot), B(\cdot))$-unreachable characteristic multiplier. Moreover, such a stabilizing solution $P_S(\cdot)$ is unique and coincides with the maximal one. ■

Methods for the numerical computation of the maximal solution include direct integration, quasilinearization (Bittanti *et al.*, 1988a) and time-invariant reformulation (Bittanti *et al.*, 1990). A comparison between the merits of these different approaches can be found in Bittanti *et al.* (1988b).

Later on, we will resort to the following theorem, the proof of which, given in Bittanti and De Nicolao (1992b), is here omitted for the sake of conciseness.

Theorem 2. Assume that the DPRE (6) admits a maximal solution $P^+(\cdot)$. Then, $P(\cdot)$ being any symmetric periodic solution of the DPRE (6), $H(t)P(t)H(t)' + R(t) > 0$, $\forall t$. ■

4. *Periodic innovations representation of cyclostationary processes*

Suppose that $F(\cdot)$ is asymptotically stable. Then, system (2) defines a unique cyclostationary process $y(\cdot)$, and can be seen as a state-space representation of such a process. We will now introduce a family of alternative state-space representations of the following type

$$\hat{x}(t+1) = F(t)\hat{x}(t) + L(t)\varepsilon(t), \tag{9a}$$

$$y(t) = H(t)\hat{x}(t) + \varepsilon(t), \tag{9b}$$

where $\varepsilon(t)$ is a zero-mean white noise with a periodic covariance matrix $\tilde{R}(t) = \tilde{R}(t+T)$. Moreover, $L(t)$ is a T-periodic matrix. If, for some $\tilde{R}(\cdot)$ and $L(\cdot)$, the cyclostationary process defined as the output of (9) coincides with the cyclostationary process obtained as the output of (2), then (9) will be said to be an innovations-like (IL) representation of the process. If, in addition, the zeros of (9) belong to the open unit disk, such a representation will be termed innovations representation. In this section, it will be shown that there exists a one-to-one correspondence between the symmetric periodic solutions of the DPRE (6) and the IL representations (9).

Theorem 3. Suppose that $F(\cdot)$ is stable and $(F(\cdot), H(\cdot))$ is observable. Consider any symmetric periodic solution $P(\cdot)$ of the DPRE (6), with the associated gain $K(\cdot)$ defined by (8b). Then, letting $L(t) = K(t)$ and $\tilde{R}(t) = R(t) + H(t)P(t)H(t)'$, equation (9) provides an IL representation of the cyclostationary process $y(\cdot)$ defined by (2). Moreover, different solutions of the DPRE lead to different IL representations.

Proof. The observability notion is output-feedback invariant. Hence, the observability of $(F(\cdot), H(\cdot))$ is equivalent to the observability of $(A(\cdot), H(\cdot))$. Then, Theorem 1.1 implies the existence of a maximal symmetric periodic solution $P^+(\cdot)$, so that, by Theorem 2, $R(t) + H(t)P(t)H(t)' > 0$, $\forall t$. Thus, $\tilde{R}(t)$ can indeed be interpreted as a covariance matrix.

Consider now the covariance matrix function $\Lambda(t, \tau) = E[y(t)y(\tau)']$ of the cyclostationary process defined by (2). This function is given by

$$\Lambda(t, \tau) = \begin{cases} H(t)\Psi_F(t, \tau+1)[F(\tau)X(\tau)H(\tau)' + G(\tau)S(\tau)], & t > \tau, \\ H(t)X(t)H(t)' + R(t), & t = \tau, \end{cases} \tag{10}$$

where the covariance matrix of the state, $X(t) = E[x(t)x(t)']$, is the unique symmetric periodic solution of the Lyapunov equation:

$$X(t+1) = F(t)X(t)F(t)' + G(t)Q(t)G(t)'. \tag{11}$$

On the other hand, the IL representation (9) defines a unique cyclostationary process too. Its covariance matrix is given by

$$\tilde{\Lambda}(t, \tau) = \begin{cases} H(t)\Psi_F(t, \tau+1)[F(\tau)\tilde{X}(\tau)H(\tau)' + L(\tau)\tilde{R}(\tau)], & t > \tau, \\ H(t)\tilde{X}(t)H(t)' + \tilde{R}(t), & t = \tau, \end{cases} \tag{12}$$

where $\tilde{X}(t) = E[\hat{x}(t)\hat{x}(t)']$ is the unique symmetric periodic solution of:

$$\tilde{X}(t+1) = F(t)\tilde{X}(t)F(t)' + L(t)\tilde{R}(t)L(t)'. \tag{13}$$

Now, by subtracting (6a) from (11), it is easily seen that $\tilde{X}(t) = X(t) - P(t)$. Consequently, after some computations, it follows that $\tilde{\Lambda}(t, \tau) = \Lambda(t, \tau)$.

Finally, it is proven that two different solutions $P_1(\cdot)$ and $P_2(\cdot)$ of the DPRE cannot give rise to the same IL representation (9). Indeed, by contradiction, assume that

there exist two symmetric periodic solutions $P_1(\cdot)$ and $P_2(\cdot)$ of the DPRE such that $\bar{R}_1(t) = \bar{R}_2(t)$ and $K_1(\cdot) = K_2(\cdot)$, where $\bar{R}_i(t) = R(t) + H(t)P_i(t)H(t)'$ and $K_i(t) = [G(t)S(t) + A(t)P_i(t)H(t)'][H(t)P_i(t)H(t)' + R(t)]^{-1}$, $i = 1, 2$. Then, let $\bar{X}_i(\cdot)$ be the unique symmetric periodic solution of the Lyapunov equation

$$\bar{X}_i(t+1) = F(t)\bar{X}_i(t)F(t)' + K_i(t)\bar{R}_i(t)K_i(t)', \; i = 1, 2.$$

Since the known terms of these two equations do coincide, it follows that $\bar{X}_1(t) = \bar{X}_2(t)$. On the other hand, $P_i(t) = X(t) - \bar{X}_i(t)$, so that $P_1(\cdot)$ and $P_2(\cdot)$ must in fact coincide too.

Conversely, once the cyclostationary process satisfying (2) is given an IL representation, one can wonder whether there exists a symmetric periodic solution of the DPRE generating such a representation. The answer is affirmative, as stated precisely in the following.

Theorem 4. Suppose that $F(\cdot)$ is stable and $(F(\cdot), H(\cdot))$ is observable. Consider an IL representation (9) for the cyclostationary process generated by (2). Then, there exists a symmetric periodic solution $P(\cdot)$ of the DPRE (6) such that $L(t) = K(t)$ and $\bar{R}(t) + H(t)P(t)H(t)'$.

Proof. Let $X(t)$ and $\bar{X}(t)$ be the covariance matrices of the state $x(t)$ of the original representation and the state $\bar{x}(t)$ of the IL representation, respectively. By assumption, the associated output covariance matrices $\Lambda(t, \tau)$ and $\bar{\Lambda}(t, \tau)$ do coincide. By comparing their expressions, (10) and (12), for $t = \tau, \tau + 1, \ldots, \tau + NT$, it follows that

$$\bar{R}(t) = R(t) + H(t)[X(t) - \bar{X}(t)]H(t)'$$

$$V(\tau+1)[F(\tau)X(\tau)H(\tau)' + G(\tau)S(\tau)] = V(\tau+1)[F(\tau)\bar{X}(\tau)H(\tau)' + L(\tau)\bar{R}(\tau)].$$

In view of the observability of $(F(\cdot), H(\cdot))$, the Observability Criterion (Section 2) guarantees that $V(\tau+1)$ is full-rank, $\forall\tau$. Therefore,

$$L(t) = [G(t)S(t) + F(t)(X(t) - \bar{X}(t))H(t)']\bar{R}(t)^{-1}.$$

By substituting the expressions for $\bar{R}(t)$ and $L(t)$ in equation (13), the conclusion is drawn that $X(\cdot) - \bar{X}(\cdot)$ satisfies the DPRE (6). Consequently, $\bar{R}(t)$ and $L(t)$ can be seen as generated by a solution of the DPRE, i.e. $\bar{R}(t) = R(t) + H(t)P(t)H(t)'$ and $L(t) = [G(t)S(t) + F(t)P(t)H(t)']\bar{R}(t)^{-1}$ with $P(t) = X(t) - \bar{X}(t)$. ■

From Theorems 3 and 4, one can conclude that, under the observability assumption, there is a one-to-one correspondence between the symmetric periodic solutions of the DPRE and the IL representation of a given cyclostationary process.

As in the stationary case, among the various IL representations, a major role is played by the innovations representation, which is characterized by the property that all its zeros belong to the open unit disk. As shown below, this is equivalent to requiring that the associated symmetric periodic solution of the DPRE is stabilizing. Indeed, an IL representation can be rewritten as

$$\hat{x}(t+1) = \hat{A}(t)\hat{x}(t) + K(t)y(t), \tag{14a}$$

$$y(t) = H(t)\hat{x}(t) + \varepsilon(t), \tag{14b}$$

where $\hat{A}(t) = F(t) - K(t)H(t)$. As seen in Lemma 1, a periodic output feedback does not affect the zeros of a periodic system. Therefore, the zeros of the IL representation (14) coincide with the zeros from $\varepsilon(\cdot)$ to $y(\cdot)$ of the system:

$$x(t+1) = \hat{A}(t)x(t),$$

$$y(t) = H(t)x(t) + \varepsilon(t).$$

According to the definition of zeros of a periodic system given in Section 2.2, such zeros turn out to be the characteristic multipliers of $\hat{A}(\cdot)$. Since $\hat{A}(\cdot)$ is the closed-loop matrix, the conclusion is that the innovations representation exists if and only if the DPRE admits a stabilizing periodic solution.

5. *PARMA models: minimum-phase condition and innovations representation*

In this section the notions of poles and zeros of a PARMA model are introduced. Moreover, the problem of the existence of a canonical (minimum-phase) PARMA representation of a cyclostationary process is discussed. This is done by resorting to a state-space representation of the original PARMA. The input–output model (1) is equivalent (i.e. has the same impulse response) to the following linear periodic system

$$x(t+1) = F(t)x(t) + G(t)e(t), \tag{15a}$$

$$y(t) = H(t)x(t) + e(t), \tag{15b}$$

where

$$F(t) = \begin{bmatrix} 0 & 0 & \cdots & 0 & a_n(t) \\ 1 & 0 & \cdots & 0 & a_{n-1}(t) \\ 0 & 1 & \cdots & 0 & a_{n-2}(t) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & a_1(t) \end{bmatrix} \tag{16a}$$

$$t) = \begin{bmatrix} a_n(t) + c_n(t) \\ a_{n-1}(t) + c_{n-1}(t) \\ a_{n-2}(t) + c_{n-2}(t) \\ \cdots \\ c_1(t) + c_1(t) \end{bmatrix}, \tag{16b}$$

Correspondingly, we will say that p is a pole of the PARMA (1) if it is a characteristic multiplier of $F(\cdot)$. As for the zeros, in view of Lemma 1, the zeros of (15) coincide with the zeros of the system

$$x(t+1) = A(t)x(t), \tag{17a}$$

$$y(t) = H(t)x(t) + e(t). \tag{17b}$$

where $A(t) = F(t) - G(t)H(t)$ is given by

$$A(t) = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_n(t) \\ 1 & 0 & \cdots & 0 & -c_{n-1}(t) \\ 0 & 1 & \cdots & 0 & -c_{n-2}(t) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -c_1(t) \end{bmatrix}. \tag{18}$$

According to Definition 1, such zeros turn out to be the characteristic multipliers of $A(\cdot)$, thus leading to

Definition 3. z is said to be a zero of the PARMA (1) if it is a characteristic multiplier of $A(\cdot)$. ■

It is immediately verified that, in the particular case of a time-invariant ARMA, such definitions coincide with the classical notions of poles and zeros. As a natural consequence of Definition 3, we will say that the PARMA model (1) is minimum-phase, if $A(\cdot)$ is stable.

Note that the state-space representation (15) of the original PARMA model (1) coincides with system (2), provided that $F(t)$, $G(t)$, $H(t)$ are defined according to (16). As for the variances $Q(t)$, $R(t)$, and the cross-covariance $S(t)$ of the disturbances $v(\cdot)$ and $w(\cdot)$ appearing in (2), one has to take $Q(t) = S(t) = R(t)$, where $R(t)$ is the variance of $e(t)$. As a consequence, matrix $B(t)$ defined in (7b) turns out to be a zero matrix $\forall t$, so that the pair $(A(\cdot), B(\cdot))$ is completely unreachable. Conversely, by means of the Observability Criterion (Section 2.1), the pair $(A(\cdot), H(\cdot))$ turns out to be completely observable.

Since $B(t) = 0$, $\forall t$, the DPRE relative to the state-space representation (15) of the PARMA is given by

$$P(t+1) = A(t)P(t)A(t)' - A(t)P(t)H(t)'[H(t)P(t)H(t)' + R(t)]^{-1}H(t)P(t)A(t)'. \tag{19}$$

We now consider the case when the PARMA is stable, i.e. all its poles lie inside the unit circle. Then, as already observed in Section 2.1, the output of the PARMA converges to a cyclostationary process: we will say that the PARMA model is an input–output representation of such a

cyclostationary process. Now, one may wonder whether, as in the stationary case, there are several PARMA representations of the same cyclostationary process and whether a canonical (minimum-phase) representation exists. The answer to these questions can be found in the following.

Theorem 5 (*Periodic spectral factorization theorem*). Consider a stable PARMA and the associated cyclostationary process $y(\cdot)$. Given any symmetric periodic solution $P(\cdot)$ of the DPRE (19), let $\tilde{c}_i(t) = c_i(t) + \gamma_{n-i+1}(t)$, where $\gamma_i(t)$ is the ith entry of the Kalman gain $\Gamma(t)$ defined in (8c). Then, the PARMA

$$y(t) = \sum_{i=1}^{n} a_i(t-i)y(t-i) + \sum_{i=1}^{n} c_i(t-i)\tilde{e}(t-i) + \tilde{e}(t), \quad (20)$$

where $\tilde{e}(\cdot) \sim WGN\ (0, R(\cdot) + H(\cdot)P(\cdot)H(\cdot)')$. provides a representation of the cyclostationary process $y(\cdot)$.

Furthermore, there exists a minimum-phase representation iff the additional assumption is made that the PARMA (1) has no unit-modulus zero. Then, the minimum-phase representation is unique and is obtained in correspondence of the unique stabilizing solution $P_S(\cdot)$ of the DPRE.

Proof. Consider the state-space reresentation (15) of the PARMA. In view of Theorem 3, letting $\tilde{R}(\cdot) = R(\cdot) + H(\cdot)P(\cdot)H(\cdot)'$ and $L(\cdot) = K(\cdot)$, system (9) provides an IL representation of the cyclostationary process $y(\cdot)$ Note, that, since $S(t) = R(t)$, from (7a), (8a) it follows that $K(t) = G(t) + \Gamma(t)$. Recall that (1) is the input–output representation of the triple $(F(\cdot), G(\cdot), H(\cdot))$ given in (16): by comparing the triples $(F(\cdot), G(\cdot), H(\cdot))$ and $(F(\cdot), G(\cdot) + \Gamma(\cdot), H(\cdot))$, it is easy to see that (20) is in fact the input–output representation of the IL system (9).

Finally, the zeros of the PARMA (20) coincide with the characteristic multipliers of $\hat{A}(\cdot)$ $F(\cdot) - K(\cdot)H(\cdot) = A(\cdot) - \Gamma(\cdot)H(\cdot)$. Therefore, the canonical minimum-phase representation is obtained in correspondence with the stabilizing solution $P_S(\cdot)$. Since $(A(\cdot), H(\cdot))$ is completely observable and $(A(\cdot), B(\cdot))$ completely unreachable, Theorem 1.3 implies that such a stabilizing solution exists iff $A(\cdot)$ has no unit-modulus characteristic multiplier.

6. *PARMA models: optimal prediction*

This section is devoted to the solution of the optimal one-step-ahead prediction problem for PARMA models. Our approach involves the application of Kalman prediction theory to the state-space representation (15) in order to derive the optimal predictor for the PARMA (1).

It is well known that the optimal periodic predictors associated with (15) can be expressed as follows

$$\begin{aligned} x(t+1/t) &= F(t)x(t/t-1) + K(t)[y(t) - y(t/t-1)] \\ &= A(t)x(t/t-1) + \Gamma(t)[y(t) \\ &\quad - y(t/t-1)] + G(t)y(t), \qquad (21a) \end{aligned}$$

$$y(t/t-1) = H(t)x(t/t-1), \qquad (21b)$$

where $\Gamma(t)[K(t)]$ conforms to (8c)((8b)), in which $P(\cdot)$ is an SPPS solution of the DPRE (19).

We are now in a position to solve the optimal prediction problem in the minimum-phase case (Theorem 6) as well as in the nonminimum-phase one (Theorem 7).

Theorem 6. Consider the PARMA given in (1) and assume that it is minimum-phase. Then, the optimal periodic steady-state predictor is

$$y(t/t-1) = -\sum_{i=1}^{n} c_i(t-i)y(t-i/t-i-1) + \sum_{i=1}^{n} [c_i(t-i) + a_i(t-i)]y(t-i). \quad (22)$$

Moreover, such a predictor is stable.

Proof. The minimum-phase assumption entails that $A(\cdot)$ is asymptotically stable and therefore system (17) is stabilizable and detectable. Consequently, the unique periodic solution of the DPRE (Theorem 1.2) is the trivial one, $P(t) = 0$, $\forall t$, which is also stabilizing. It is a matter of simple computations to derive (22) from (21). The stability of (22) follows directly from the stability of $A(\cdot)$. As a matter of fact, the zeros of a minimum-phase PARMA become the poles of the associated optimal predictor. ■

If the minimum-phase condition is not met with, $A(\cdot)$ is not stable and the uniqueness of the SPPS solution of the DPRE (19) is no more guaranteed. Depending on the initial covariance of the state of system (17), there are, in general, several optimal periodic predictors, each of which corresponds to an SPPS solution of the DPRE. Among all the predictors it is then wise to choose the stabilizing one, if any. This motivates the following theorem.

Theorem 7. Consider a PARMA system with no unit-modulus zero. Then, the stable optimal predictor is given by

$$y(t/t-1) = -\sum_{i=1}^{n} \tilde{c}_i(t-i)y(t-i/t-i-1) + \sum_{i=1}^{n} [\tilde{c}_i(t-i) + a_i(t-i)]y(t), \quad (23)$$

where $\tilde{c}_i(t) = c_i(t) + \gamma_{n-i+1}(t)$, $\gamma_i(t)$ being the ith entry of the Kalman gain $\Gamma(t)$ associated with the stabilizing solution $P_S(\cdot)$ of the DPRE (19).

Proof. By assumption $A(\cdot)$ has no unit-modulus characteristic multiplier. Moreover, we know that the pair $(A(\cdot), H(\cdot))$ is observable. We can then apply Theorem 1.3 to conclude that the DPRE (19) admits a stabilizing solution, which is positive semidefinite. Consider now the corresponding optimal periodic predictor (21). The PARMA (23) is in fact the input–output representation of such a predictor. The stability of (23) stems from the stability of the closed loop matrix $\hat{A}(\cdot) = A(\cdot) - \Gamma(\cdot)H(\cdot)$. ■

From Theorem 5, 6 and 7, it is apparent that solving the prediction problem for a stable nonminimum-phase PARMA is equivalent to finding a minimum-phase PARMA realization of the associated cyclostationary process $y(\cdot)$ and then applying (22) to such a "canonical" PARMA.

7. *Conclusions*

In this paper, the factorization problem for cyclostationary processes and the optimal prediction problem for ARMA models with periodic coefficients (PARMA models) have been studied. It has been shown that there is a one-to-one correspondence between periodic factorizations and periodic solutions of a difference periodic Riccati equation. On the basis of such a result, the prediction problem for PARMA models has been given a solution that can be outlined as follows

(a) Check the minimum-phase condition by verifying whether the zeros of the PARMA system are stable, i.e. whether the multipliers of $A(\cdot)$ given in (18) belong to the open unit disk.
(b) If in the affirmative, the optimal one-step-ahead predictor is given by (22)
(c) Otherwise, if some zeros of the PARMA lie on the unit circle, there is no stabilizing optimal periodic predictor.
(d) If no zero lies on the unit circle, compute the stabilizing solution $P_S(\cdot)$ of the periodic Riccati equation (19) and the corresponding Kalman gain $\Gamma(\cdot)$. Then, the optimal predictor is given by (23)

Note that, if the original PARMA is stable, this last step can be interpreted as applying the standard prediction rule (22) to its canonical minimum-phase representation.

Acknowledgements—This paper has been supported by MURST—project "models identification, systems control, signal processing"—and Centro di Teoria dei Sistemi of the CNR, Milano, Italy.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Anderson, B. D. O., J. B. Moore and S. G. Loo (1969) Spectral factorization of time-varying covariance functions. *IEEE Trans. on Information Theory*, **IT-15**, 550–557.

Anderson, B. D. O. and P. J. Moylan (1974). Spectral factorization of a finite-dimensional nonstationary matrix covariance. *IEEE Trans. Aut. Control*, **AC-19**, 680–692.

Bittanti, S. (1986). Deterministic and stochastic linear periodic systems. In S. Bittanti (Ed.), *Time Series and Linear Systems*. Springer, Berlin.

Bittanti, S. and P. Bolzern (1985). Discrete-time linear periodic systems: Gramian and modal criteria for reachability and controllability. *Int. J. Control*, **41**, 909–928.

Bittanti, S., P. Colaneri and G. De Nicolao (1986). Discrete-time linear periodic systems: a note on the reachability and controllability interval length. *Syst. & Control Letters*, **8**, 75–78.

Bittanti, S., P. Colaneri and G. De Nicolao (1988a). The difference periodic Riccati equation for the periodic prediction problem. *IEEE Trans. Aut. Control*, **AC-33**, 706–712.

Bittanti, S., P. Colaneri and G. De Nicolao (1988b). Two techniques for the solution of the discrete-time periodic Riccati equation. *Proc. Conf. on Linear Algebra in Signals, Systems and Control*, Boston, U.S.A. 12–14 August 1986; SIAM; Philadelphia, pp. 315–331.

Bittanti, S., P. Colaneri and G. De Nicolao (1990). An algebraic Riccati equation for the discrete-time periodic prediction problem. *Syst. & Control Letters*, **14**, 71–78.

Bittanti, S., P. Colaneri and G. De Nicolao (1991). The periodic Riccati equation. In S. Bittanti, A. J. Laub and J. C. Willems (Eds), *The Riccati Equation*. Springer, Berlin, pp. 127–162.

Bittanti, S. and G. De Nicolao (1990). Periodic ARMA models: Prediction and minimum-phase condition. *Preprints 11th IFAC World Congress*, August 1990, Tallinn, Vol. 6, pp. 200–205.

Bittanti, S. and G. De Nicolao (1992a). Spectral factorization of linear periodic systems with application to the optimal prediction of periodic ARMA models. *Int Rep.* 92-002. Centro Teoria dei Sistemi del CNR, Dipartimento di Elettronica del Politecnico di Milano, Milano.

Bittanti, S. and G. De Nicolao (1992b). A note on the periodic Riccati equation and innovations representations of cyclostationary processes. *Proc. 31st Conf. on Decision and Control*, Dec. 1992, Tucson, U.S.A.

Bolzern, P., P. Colaneri and R. Scattolini (1986). Zeros of discrete-time linear periodic systems. *IEEE Trans. Aut. Control*, **AC-31**, 1057–1058.

de Souza, C. E. (1989a). Existence conditions and properties for the maximal periodic solution of periodic Riccati difference equations. *Int. J. Control*, **50**, 731–742.

de Souza, C. E. (1989b). Maximal periodic solutions of periodic Riccati difference equations. *Proc. Workshop on the Riccati Equation in Control, Systems and Signals*, Como, Italy, pp. 177–182.

de Souza, C. E. (1991). Periodic strong solution for the optimal filtering problem of linear discrete-time periodic systems. *IEEE Trans. Aut. Control*, **AC-36**, 333–338.

Gardner, W. A. (1991). Exploitation of spectral redundancy in cyclostationary signals *IEEE Signal Processing Magazine*, **8**, 14–36.

Gardner, W. A. and L. E. Franks (1975). Characterization of cyclostationary random processes *IEEE Trans. Inform. Theory*, **IT-21**, 4–14.

Grasselli, O. M. and S Longhi (1988) Zeros and poles of linear periodic multivariate discrete-time systems. *Circuits Systems Signal Processing*, **7**, 361–380.

Jones, R. H. and W. M. Brelsford (1967). Time series with periodic structure. *Biometrika*, **54**, 403–408.

Kailath, T. (1969). Fredholm resolvents, Wiener–Hopf equations, and Riccati differential equations *IEEE Trans. Information Theory*, **IT-15**, 665–672.

Lancaster, P., A. C. M. Ran and L. Rodman (1986). Hermitian solutions of the discrete algebraic Riccati equation. *Int J. Control*, **44**, 777–802

Pagano, M. (1978) On periodic and multiple autoregressions *The Annals of Statistics*, **6**, 1310–1317

Vecchia, A V. (1985) Maximum likelihood estimation for periodic autoregressive moving average models *Technometrics*, **27**, 375 384

Automatica, Vol. 29, No. 2, pp. 523–526, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Steady-state Errors in Discrete-time Control Systems*

IRWIN W. SANDBERG† and LILIAN Y. XU†

Key Words—Control theory; discrete nonlinear systems; dynamic response; dynamic stability; feedback control; iterative methods; nonlinear control systems; nonlinear systems; stability criteria; system analysis.

Abstract—We address discrete-time control systems containing a memoryless nonlinear element and characterize the error response to certain inputs when the circle condition for stability is met.

1. *Introduction*

THE CONCEPT of steady-state errors often plays an important role in the design of control systems (see, for example, Kuo (1991)). For the discrete-time case, attention is frequently focused on the linear unity-feedback system of Fig. 1, and on its error response e to certain standard inputs r such as step functions, ramps, etc. For such systems, it is a classical result that the number of poles of G at the point $(1 + j0)$, often referred to as the "system type", determines whether the system error is unbounded, approaches a nonzero constant, or approaches zero.

2. *Results for nonlinear systems*

In this paper we report on an extension of the theory for linear systems described above. The extension concerns control systems of the form shown in Fig 2, where N takes into account a memoryless nonlinear element of the sector type. We show that much more general results along the lines of those for the classical linear case hold. This involves a hypothesis analogous to the circle condition (Sandberg, 1964) for the stability of continuous-time systems. Our results are along the lines of those in Sandberg and Johnson (1993) where only continuous-time systems are treated. Throughout the paper when we say that a limit exists, we mean that it exists as a real number.

For the systems we consider, an additional degree of complexity arises concerning the evaluation of the limit of the steady-state response when it exists, because the systems are nonlinear. We give a convergent algorithm for computing the limit whenever it exists.

It will become clear that our inputs are more general than merely steps, ramps, etc. We consider more general inputs to establish that under our conditions the system response possesses a degree of robustness with respect to inputs. Additional comments are given in Section 3.1.

2.1. *The nonlinear system and assumptions.* Let $\mathscr{Z}_+$ be the set of nonnegative integers, and let S denote the set of real-valued functions defined on $\mathscr{Z}_+$.

Referring to Fig. 2, it is assumed throughout that r, e, v, and y belong to S, and that y and v are related by

$$y_t = (Lv)_t + d_t, \quad t \in \mathscr{Z}_+,$$

where the operator L is defined by

$$(Lx)_t = \sum_{\tau=0}^{t} g_{t-\tau} x_\tau, \quad t \in \mathscr{Z}_+,$$

for $x \in S$. Here $g \in S$ is the inverse z-transform of

$$G(z) = \frac{p(z)}{(z-1)^{\rho} q(z)},$$

where $\rho \in \mathscr{Z}_+$ is the number of poles of G at $z = 1 + j0$ and p and $(z-1)^{\rho} q$ are relatively prime real polynomials such that $\deg(p) \leq \deg((z-1)^{\rho} q)$. As usual, $(z-1)^0$ means 1. The function $d \in S$ takes into account initial conditions, and it is assumed that d is the inverse z-transform of

$$D(z) = \frac{m(z)}{(z-1)^{\rho} q(z)},$$

where m is a real polynomial such that D is proper. With regard to the nonlinear portion of the system, we suppose that $(Nx)_t = \eta[x_t]$, $t \in \mathscr{Z}_+$ for $x \in S$ where η maps $\mathscr{R}$ into $\mathscr{R}$ such that $\eta(0) = 0$ and there are two real constants α and β such that

$$\alpha \leq \frac{\eta(a) - \eta(b)}{a - b} \leq \beta,$$

for $a \neq b$. Define $c_0 \in \mathscr{R}$ by

$$c_0 = \tfrac{1}{2}(\alpha + \beta),$$

and assume throughout that $c_0 \neq 0$.

The system under consideration is governed by

$$r_t - d_t = e_t + \sum g_{t-\tau} \eta[e_\tau], \quad t \in \mathscr{Z}_+. \tag{1}$$

We assume that the following three conditions are met, and will refer to this assumption as B.1:

(i) $1 + c_0 g_0 \neq 0$,
(ii) $1 + c_0 G(z) \neq 0$ for $|z| \geq 1$, and
(iii) $\frac{1}{2}(\beta - \alpha) \sup_{|z|=1} |H(z)| < 1$ where H denotes the rational function defined by

$$H(z) = \frac{G(z)}{1 + c_0 G(z)},$$

at all points z at which G is regular. (By (i) and (ii) this function is proper and has no poles on or outside the unit circle. We will use the fact that H is the z-transform of an $\ell_1(\mathscr{Z}_+)$ function h.‡)

Hypothesis B.1 is analogous to the key condition A.1 of Sandberg and Johnson (1990) which is stated in circle-criterion form; B.1 too can be expressed in that form.

* Received 28 August 1991; revised 23 April 1992; received in final form 14 May 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor R. V. Patel under the direction of Editor H. Kwakernaak.

† Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, U.S.A.

‡ Of course, $l_1(\mathscr{Z}_+) = \{s \in S : \sum_n |s_n| < \infty\}$.

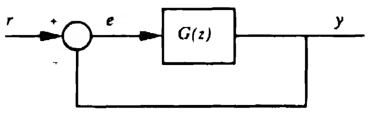


FIG. 1. Linear control system.

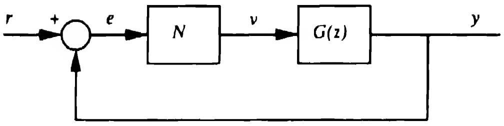


FIG. 2. Nonlinear control system.

If $g_0 = 0$, e_t in (1) is defined explicitly for each t, and so it is clear that for each $r \in S$ there is a unique $e \in S$ such that (1) is satisfied. It can be shown that this existence and uniqueness holds even if $g_0 \neq 0$.*

Define $\bar{\eta}: \mathscr{R} \to \mathscr{R}$ by $\bar{\eta}(a) = \eta(a) - c_0 a$, $a \in \mathscr{R}$. The following proposition plays a central role in the statement of Theorem 1 our main result which appears below.

Proposition 1. Suppose that the assumptions in Section 2.1 are met. Then for each $\gamma \in \mathscr{R}$, there exists a unique $\xi \in \mathscr{R}$ such that

$$\gamma = \xi + H(1)\bar{\eta}(\xi),$$

and one has

$$\xi = \lim_{k \to \infty} \xi_k,$$

where $\xi_{k+1} = \gamma - H(1)\bar{\eta}(\xi_k)$ and $\xi_0 \in \mathscr{R}$ is arbitrary. Also, $c \triangleq \frac{1}{2}(\beta - \alpha)\,|H(1)| < 1$ and

$$|\xi - |\xi_k| \leq \frac{c^k}{1-c}|\xi_0 - \gamma + H(1)\bar{\eta}(\xi_0)|, \quad k \geq 1.$$

This follows from the contraction-mapping fixed-point theorem and its proof (Kolmogorov and Fomin, 1957) using the inequality (see B.1) $\frac{1}{2}(\beta - \alpha)\,|H(1)| < 1$ Referring to the proposition, let $\Theta: \mathscr{R} \to \mathscr{R}$ be the map that takes γ into ξ. Our result, which concerns the relationship between the nature of e and the system type, is the following.

Theorem 1. Assuming that the assumptions described in Section 2.1 are satisfied:

(i) If r approaches a limit l as $t \to \infty$, then $\sigma \triangleq \lim_{t \to \infty} e_t$ exists. Moreover, $\sigma \neq 0$ if and only if $l \neq 0$ and $\rho = 0$, and then $\sigma = \Theta\{l[1 + c_0 G(1)]^{-1}\}$.

(ii) Assuming that $r_t = \sum_{j=0}^{v} a_j t^j$ for $t \in \mathscr{T}_+$, where the a_j are real, v is a positive integer, and $a_v \neq 0$,
(a) e is unbounded if $v > \rho$, and
(b) if $v \leq \rho$, then e approaches a limit as $t \to \infty$. This limit is $\Theta[a_v v!\, q(1)/c_0 p(1)]$ if $v = \rho$ and zero otherwise.†

* In this connection, $g_0 = 0$ in the important case in which $G(z)$ represents a sampled system consisting of a zero-order hold followed by a strictly-proper rational transfer function Q. This follows from $g_0 = \lim_{|z| \to \infty} G(z)$, the initial-value theorem for Laplace transforms, and the standard relation $G(z) = (1 - z^{-1})Z\{s^{-1}Q(s)\}$, where $Z\{s^{-1}Q(s)\}$ denotes the z-transform of the inverse Laplace transform of $s^{-1}Q(s)$.

† Referring to the argument of Θ in (ii) of Theorem 1, for pedagogical reasons some readers may wish to observe that $c_0[p(1)/q(1)]$ is simply the product of the "average gain" $\frac{1}{2}(\alpha + \beta)$ of N times the coefficient of the $(z-1)^{-\rho}$ term in the partial-fraction expansion of G.

3. *Proof of Theorem* 1

Assume that the hypotheses of the theorem are met, and let I denote the identity map on S. We will use the following lemma.

Lemma 1. The inverse of $(I + c_0 L): S \to S$ exists and has the representation

$$[(I + c_0 L)^{-1}x]_t = x_t - c_0 \sum_{\tau=0}^{t} h_{t-\tau} x_\tau, \quad t \in \mathscr{T}_+,$$

for $x \in S$ (where as indicated earlier h is the inverse z-transform of H).

Proof of Lemma 1

It is clear that g and h are related by

$$g_t - h_t = c_0 \sum_{\tau=0}^{t} g_{t-\tau} h_\tau, \quad t \in \mathscr{T}_+. \tag{2}$$

Using (2) one finds that $y \in S$ defined by

$$y_t = x_t - c_0 \sum_{\tau=0}^{t} h_{t-\tau} x_\tau, \quad t \in \mathscr{T}_+, \tag{3}$$

satisfies

$$y_t + c_0 \sum_{\tau=0}^{t} g_{t-\tau} y_\tau = x_t, \quad t \in \mathscr{T}_+, \tag{4}$$

for $x \in S$. On the other hand, using (2), we find that if $y \in S$ satisfies (4) for some $x \in S$, then y is given by (3) Thus $(I + c_0 L)$ is one-to-one and onto, which completes the proof

By Lemma 1, (1) is equivalent to

$$(I + c_0 L)^{-1}(r - d) = e + (I + c_0 L)^{-1} L\bar{N}e, \tag{5}$$

for $r \in S$, where $(\bar{N}e)_t = \bar{\eta}[e_t]$ for all $t \in \mathscr{T}_+$, and

$$[(I + c_0 L)^{-1} Lx]_t = \sum_{\tau=0}^{t} h_{t-\tau} x_\tau, \quad t \in \mathscr{T}_+,$$

for $x \in S$.

The z-transform of $\delta_1 \triangleq (I + c_0 L)^{-1} d$, which appears on the left side of (5), exists, is proper, and is

$$\frac{m(z)}{(z-1)^{\rho} q(z) + c_0 p(z)} \tag{6}$$

By B.1, (6) has no poles on or outside the unit circle So, $\delta_{1t} \to 0$ as $t \to \infty$. Therefore

$$-\delta_{1t} + [(I + c_0 L)^{-1} r]_t = e_t + \sum_{\tau=0}^{t} h_{t-\tau}\bar{\eta}(e_\tau), \quad t \in \mathscr{T}_+, \tag{7}$$

for $r \in S$, where δ_1 is as described

Proof of Part (i)

We need two lemmas.

Lemma 2. If $r_t \to l \in \mathscr{R}$ as $t \to \infty$, then $[(I + c_0 L)^{-1} r]_t$ approaches

$$l\left(1 - c_0 \sum_{\tau=0}^{\infty} h_\tau\right).$$

This follows from Lemma 1 and a version of the final-value theorem for z-transforms.

Lemma 3. Let $\gamma \in \mathscr{R}$, and let $\delta \in S$ be such that $\delta_t \to 0$ as $t \to \infty$ If

$$\delta_t + \gamma = e_t + \sum_{\tau=0}^{t} h_{t-\tau}\bar{\eta}(e_\tau), \quad t \in \mathscr{T}_+, \tag{8}$$

then $e_t \to \Theta(\gamma)$ as $t \to \infty$.

Proof of Lemma 3 Let $\xi = \Theta(\gamma)$. Then

$$\gamma = \xi + \sum_{\tau=0}^{\infty} h_\tau \bar{\eta}(\xi), \quad t \in \mathscr{T}_+,$$

and so

$$\gamma - \sum_{\tau=t+1}^{\infty} h_\tau \bar{\eta}(\xi) = \xi + \sum_{\tau=0}^{t} h_{t-\tau}\bar{\eta}(\xi), \quad t \in \mathscr{T}_+. \tag{9}$$

Since the difference of the left sides of (9) and (8) approaches zero as $t \to \infty$, and B.1 holds, by Theorem 2′ of Sandberg (1965), one has $e_t \to \Theta(\gamma)$ as $t \to \infty$, as claimed.

We shall use Proposition 1 and the observation that $\bar{\eta}(0) = 0$ to prove the remainder of Part (i).

Clearly $l = 0$ implies $\sigma = 0$. If $\rho > 0$, $p(1) \neq 0$ and

$$H(1) = \left. \frac{p(z)}{(z-1)^\rho q(z) + c_0 p(z)} \right|_{z=1} = \frac{1}{c_0}.$$

So, then $1 - c_0 H(1) = 0$, yielding $\sigma = 0$. Therefore, $\sigma \neq 0$ implies $l \neq 0$ and $\rho = 0$.

Now suppose $l \neq 0$ and $\rho = 0$. Then $1 - c_0 H(1) = [1 + c_0 G(1)]^{-1}$ and $l[1 + c_0 G(1)]^{-1} \neq 0$, showing that $\sigma \neq 0$.

Proof of Part (ii). We begin with two definitions: let $w_j \in S$ be defined by

$$w_{jt} = t^j - c_0 \sum_{\tau=0}^{t} h_{t-\tau} \tau^j, \quad t \in \mathscr{T}_+,$$

for $j = 1, 2, \ldots, \nu$. For every positive integer k, let $x \approx t^k$ mean that $x \in S$ and there is a nonzero constant μ such that $x_t / t^k \to \mu$ as $t \to \infty$.

We will use the following lemma.

Lemma 4.

(a) $j > \rho \Rightarrow w_{jt} \approx t^{(j-\rho)}$,
(b) $j = \rho \Rightarrow w_{jt} \to j!\, q(1)/c_0 p(1)$ as $t \to \infty$, and
(c) $j < \rho \Rightarrow w_{jt} \to 0$ as $t \to \infty$.

Proof of Lemma 4. Choose j, and let W_j denote the z-transform of w_j, which clearly exists. One has

$$W_j(z) = Z\{t^j\}[1 - c_0 H(z)],$$

where $Z\{t^j\}$, the z-transform of t^j, is of the form $T_j(z)/(z-1)^{j+1}$ with $T_j(z)$ a polynomial in z of degree j with no zeros at $z = 1$. Thus, we have

$$W_j(z) = \frac{T_j(z) q(z)}{(z-1)^{j+1-\rho}[(z-1)^\rho q(z) + c_0 p(z)]}.$$

Since p and $(z-1)^\rho q$ are relatively prime, by B.1, we have

$$(z-1)^\rho q(z) + c_0 p(z) \neq 0 \quad \text{for} \quad |z| \geq 1.$$

Thus, part (a) follows from the fact that W_j has $j + 1 - \rho$ poles at $z = 1$, and part (c) from the fact that W_j has no poles at $z = 1$ when $j < \rho$. If $j = \rho$, then W_j has a simple pole at $z = 1$ and the associated residue is $T_j(1) q(1)/c_0 p(1)$. Since $T_j(1) = j!$, we have also proved part (b).

Continuing with the proof of the theorem, let r be as described in part (ii). Referring to the left side of (7),

$$(I + c_0 L)^{-1} r = \sum_{j=0}^{\nu} a_j w_j. \tag{10}$$

If $\nu > \rho$,

$$\sum_{j=0}^{\nu} a_j w_j = \sum_{j=0}^{\rho} a_j w_j + \sum_{j=\rho+1}^{\nu} a_j w_j,$$

in which, by (a), (b), and (c) of Lemma 4, the first sum on the right side is bounded and

$$\sum_{j=\rho+1}^{\nu} a_j w_j \approx t^{(\nu-\rho)}.$$

Since $h \in l_1(\mathscr{T}_+)$, e bounded implies that the right side of (7) is uniformly bounded in t. Therefore, e must be unbounded if $\nu > \rho$.

When $\nu < \rho$, the right side of (10) approaches zero as $t \to \infty$, by (c) of Lemma 4. Thus, using Lemma 3 and Proposition 1, $e_t \to 0$ as $t \to \infty$.

Finally, using (b) and (c) of Lemma 4, if $\nu = \rho$, the right side of (10) approaches $a_\nu \nu!\, q(1)/c_0 p(1)$ and, by Lemma 3, e_t approaches the limit stated in the theorem. This completes the proof of the theorem.

3.1. *Comments*. Referring to Part (ii) of Theorem 1, in classical linear theory attention is focused only on the case in which $\nu \in \{1, 2\}$ (and r is a monomial). We have considered $\nu \in \{1, 2, \ldots\}$ to illustrate the general pattern that results.

Our results are, of course, not restricted to the configuration of Fig. 2. As is well known, more general systems are governed by an equation of the form (1). A related paper is Miller *et al.* (1989) which considers the effects of quantizers on the error performance of control systems.

Cases in which a linear subsystem $K(z)$ precedes N are of interest. Two results for $K(z) = (z-1)^{-1}$, along the general lines of the proof of Theorem 1, are given in the next section. In addition, using a discrete-time analog of Wiener's tauberian theorem it is shown that, under a natural extension of the assumptions in Section 2.1, it is *not true* that the system error e approaches a limit whenever e when $\bar{\eta} = 0$ approaches a limit. This shows that certain specific extensions of the proof of Theorem 1 can be carried out, and that others cannot.

4. *Systems with an integrator preceding* N

In this section we consider the system of Fig. 3, in which the blocks labeled N and $G(z)$ are as described in Section 2.1, and r, e, x, and y are members of S with r the inverse z-transform of a proper rational function. The block labeled $(z-1)^{-1}$ represents an integrator. It introduces the relation

$$x_t = x_0 + \sum_{\tau=0}^{t-1} e_\tau, \quad t \in \{1, 2, \ldots\}. \tag{11}$$

Let B.2 stand for the condition that B.1 is met with $G(z)$ replaced with $(z-1)^{-1} G(z)$. Under B.2 it is straightforward to show that x in Fig. 3 satisfies an equation of the form

$$x_t + \sum_{\tau=0}^{t} w_{t-\tau} \bar{\eta}(x_\tau) = u_t, \quad t \in \mathscr{T}_+, \tag{12}$$

where w and u are inverse z-transforms of proper rational functions, that $w \in l_1(\mathscr{T}_+)$, and that we have

$$\tfrac{1}{2}(\beta - \alpha) \sup_{|z|=1} |W(z)| < 1,$$

where W is the z-transform of w. Assume that B.2 is satisfied.

The set S contains a unique solution x of (12). By (11), one has

$$x_{t+1} - x_t = e_t, \quad t \in \mathscr{T}_+,$$

and using (12),

$$u_{t+1} - u_t = (x_{t+1} - x_t) + \sum_{\tau=0}^{t} w_{t-\tau}[\bar{\eta}(x_{\tau+1}) - \bar{\eta}(x_\tau)] + w_{t+1} \bar{\eta}(x_0), \quad t \in \mathscr{T}_+.$$

Hence, we have

$$e_{0t} = e_t + \sum_{\tau=0}^{t} w_{t-\tau}[\bar{\eta}(x_{\tau+1}) - \bar{\eta}(x_\tau)] + w_{t+1} \bar{\eta}(x_0), \quad t \in \mathscr{T}_+, \tag{13}$$

where $e_{0t} = u_{t+1} - u_t$, $t \in \mathscr{T}_+$. By $|\bar{\eta}(a) - \bar{\eta}(b)| \leq \tfrac{1}{2}(\beta - \alpha)|a - b|$ for real a and b,

$$\bar{\eta}(x_{t+1}) - \bar{\eta}(x_t) = \varphi_t (x_{t+1} - x_t), \quad t \in \mathscr{T}_+,$$

with $|\varphi_t| \leq \tfrac{1}{2}(\beta - \alpha)$, and using (13), one has

$$e_{0t} + d_t = e_t + \sum_{\tau=0}^{t} w_{t-\tau} \varphi_\tau e_\tau, \quad t \in \mathscr{T}_+, \tag{14}$$

where $d \in S$ is given by $d_t = -w_{t+1} \bar{\eta}(x_0)$ for each t.

Below results are given related to those of Section 2. Notice that e_0 in (13) in the system error when the nonlinear

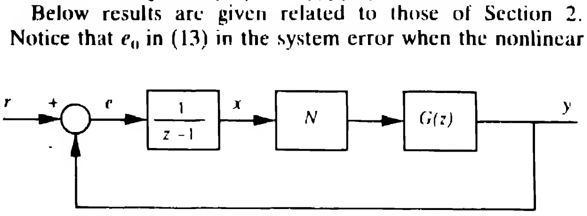


FIG. 3. Nonlinear control system.

part $\tilde{\eta}$ of η vanishes.* We show in Theorem 2 below that $e_0 \to 0$ (i.e. $e_{0t} \to 0$ as $t \to \infty$) if and only if $e \to 0$. Theorem 3 asserts that e is bounded if and only if e_0 is bounded. Observe that these results are along the lines of the proofs in Section 2. On the other hand, in contrast with the results in Section 2, Theorem 4 shows that it is not true that e_0 approaches a nonzero limit implies that e approaches a limit.

Theorem 2. Under the condition indicated, $e_0 \to 0$ if and only if $e \to 0$.

Proof of Theorem 2. If $e_0 \to 0$, then $(e_0 + d) \to 0$ and, by (14) and Theorem 2′ of Sandberg (1965), $e \to 0$.

Conversely, let $e \to 0$ and consider (14). Since $w \in l_1(\mathcal{Z}_+)$, $e \to 0$, and φ is bounded, it follows that $\sum_{\tau=0}^{t} w_{t-\tau}\varphi_\tau e_\tau \to 0$. Thus (using $d \to 0$) $e_0 \to 0$.

Theorem 3. Under the conditions indicated e_0 is bounded if and only if e is bounded.

Proof of Theorem 3. Consider (14). Since $w \in l_1(\mathcal{Z}_+)$ and $W(z)$ is rational, we have $\tau^2 w \in l_1(\mathcal{Z}_+)$ (notation of Sandberg (1965)). In addition, d is bounded and as indicated above $|\varphi_t| \le \frac{1}{2}(\beta - \alpha)$ for all t. Thus, e_0 bounded implies that e is bounded, by Theorem 2′ of Sandberg (1965).

On the other hand, suppose that e_0 is unbounded. Then by (14) the hypothesis that e is bounded leads to the contradiction that e_0 is bounded. Therefore e is unbounded. This completes the proof

Comment. With regard to explicit corresponding results in the context of Section 2, we see that if $\hat{e}_0$ denotes the left side of (7), then e in (7) is bounded if and only if $\hat{e}_0$ is bounded, and similarly e in (7) approaches a limit if and only if $\hat{e}_0$ approaches a limit.†

Returning now to the system of Fig. 3, we prove the following

Theorem 4. Under the conditions indicated, the existence of $\lim_{t\to\infty} e_{0t}$ does not imply that $\lim_{t\to\infty} e_t$ exists.

Proof of Theorem 4. We use Lemma 5, below.

Lemma 5. Assume that the conditions indicated are met, that $p(e^{-i\theta}) \neq 0$ for $\theta \in [-\pi, \pi)$ ($p(z)$ is the numerator of $G(z)$), and that the limit of $[\tilde{\eta}(x_{t+1}) - \tilde{\eta}(x_t)](x_{t+1} - x_t)^{-1}$ fails to exist for any strictly monotone sequence $x_0, x_1, \ldots,$ in $\mathcal{R}$ such that $|x_{t+1} - x_t| \le \gamma$, $t \in \mathcal{Z}_+$, for some $\gamma \in \mathcal{R}$. Assume also that $\lim_{t\to\infty} e_{0t}$ exists and is nonzero. Then $\lim_{t\to\infty} e_t$ fails to exist.

Proof of Lemma 5. Let the hypotheses hold and consider (14). Suppose (to obtain a contradiction) that $l \triangleq \lim_{t\to\infty} e_t$ exists. Then $l \neq 0$, by Theorem 2.

By (14), $\lim_{t\to\infty} \sum_{\tau=0}^{t} w_{t-\tau}\varphi_\tau e_\tau$ exists Since φ is bounded, and, by Theorem 3, e is bounded, $\varphi_\tau e_\tau$ is uniformly bounded in τ. In addition, the z-transform of w is $p(z)[(z-1)^{p+1}q(z) + c_0 p(z)]^{-1}$, showing that $W(e^{-i\theta}) \neq 0$ for $\theta \in [-\pi, \pi)$. By a general form Rudin (1962) of Wiener's tauberian theorem,

$$\lim_{\tau\to\infty} \varphi_\tau e_\tau, \tag{15}$$

exists.

By the existence of (15) and $l \neq 0$, φ_τ must have a limit at ∞. But since $l \neq 0$, x is unbounded, eventually monotone, and there is a $T \in \mathcal{Z}_+$ such that $|x_{t+1} - x_t| \le |l| + 1$ and $x_{t+1} - x_t \neq 0$ for $t \ge T$. Since

$$[\tilde{\eta}(x_{t+1}) - \tilde{\eta}(x_t)](x_{t+1} - x_t)^{-1} = \varphi_t + c_0, \quad t \ge T, \tag{16}$$

the limit of the left side of (16) must exist. This contradiction completes the proof of the lemma.

Now choose $G(z) = \frac{1}{2}$ and $\eta(a) = a + \tilde{\eta}(a)$, where $\tilde{\eta}(a)$ is odd symmetric, continuous, piecewise affine and such that

$$\eta'(a) = \begin{cases} \frac{1}{2}, & a \in (\sigma(2k-1), \sigma(2k)), \quad k \in \{1, 2, \ldots\} \\ -\frac{1}{2}, & a \in (\sigma(2k), \sigma(2k+1)), \quad k \in \mathcal{Z}_+, \end{cases}$$

where $\sigma(n) = \sum_{k=0}^{n} k$, $n \in \mathcal{Z}_+$. In this case, $\frac{1}{2} \le [\eta(a) - \eta(b)](a-b)^{-1} \le \frac{3}{2}$ for $a \neq b$, $c_0 = 1$, $1 + c_0(z-1)^{-1}G(z) \neq 0$ for $|z| \ge 1$, and $1 + c_0\tilde{g}_0 \neq 0$, where $\tilde{g}_0$ is the value at $t = 0$ of the inverse z-transform of $(z-1)^{-1}G(z)$. Also, $W(z)$ (i.e. $(z-1)^{-1}G(z)[1 + (z-1)^{-1}G(z)]^{-1}$) $= (2z-1)^{-1}$, $\sup_{|z|=1} |W(z)| = 1$, and so $\frac{1}{2}(\beta - \alpha) \sup_{|z|=1} |W(z)| < 1$, which shows that B.2 is met. Since η and p here meet the conditions of Lemma 5, the proof of Theorem 4 is complete.

5. *Conclusions*

The concept of steady-state errors often plays an important role in the design of control systems. Attention is frequently focused on the response of systems to certain standard inputs such as step functions, ramps, etc. For stable discrete-time linear systems, it is a classical result that the number of z-domain poles of the loop gain at the point $(1 + j0)$, often referred to as the "system type", determines whether the system error is unbounded, approaches a nonzero constant, or approaches zero.

In this paper, we address nonlinear control systems of the form shown in Fig. 2 and show that, when a hypothesis analagous to the circle condition for the stability of continuous-time systems is met, corresponding propositions hold concerning the relationship between the nature of the error response and the system type. A convergent algorithm is given for computing the limit of the steady-state response whenever it exists. Results are also given concerning cases in which an integrator precedes N.

Our approach can be used to obtain related propositions concerning the error response of systems with both inputs and disturbances.

References

Kolmogorov, A. N. and S. V. Fomin (1957). *Elements of the Theory of Functions and Functional Analysis* Garylock Press, Rochester, New York, p. 43

Kuo, B. C. (1991). *Automatic Control Systems*. Prentice Hall, Englewood Cliffs, NJ, pp. 380–384.

Miller, R. K., A. N. Michel and J. A. Farrell (1989). Quantizer effects on steady-state error specifications of digital feedback control systems. *IEEE Trans. Aut. Control*, **34**, 651–654.

Rudin, W. (1962). *Fourier Analysis on Groups*. Interscience, New York, p. 163.

Sandberg, I. W. (1964). A frequency domain condition for the stability of feedback systems containing a single time-varying nonlinear element. *Bell System Technical Journal*, **43**, 1601–1608.

Sandberg, I. W. (1965). On the boundedness of solutions of nonlinear integral equations. *Bell System Technical Journal*, **44**, 439–453.

Sandberg, I. W. and K. K. Johnson (1993). Steady-state errors in nonlinear control systems. *33rd Midwest Symposium on Circuits and Systems*; *IEEE Trans. Aut. Control* (to appear).

* Notice that we use e_0 for two different purposes. Here it is a function: earlier it denotes e at $t = 0$.

† Theorem 2 holds also in the setting of Section 2 (see (7) and Theorem 2′ of Sandberg (1965)).

Automatica, Vol. 29, No. 2, pp. 527–530, 1993
Printed in Great Britain

0005-1098/93 \$6.00 + 0.00

Brief Paper

On the Generic Controllability of Continuous Generalized State-space Systems*

F. N. KOUMBOULIS† and P. N. PARASKEVOPOULOS†

Key Words—Controllability; observability; structural controllability; system theory; linear systems.

Abstract—In this paper using a duality property between regular and generalized state-space systems, the necessary and sufficient conditions for the generic controllability of continuous generalized state-space systems are established. the necessary and sufficient conditions for the generic observability of such systems are also derived.

1. *Introduction*

THIS PAPER is devoted to the problem of generic controllability of continuous generalized state-space (g.s.s.) systems, i.e. of systems described by Verghese *et al.* (1981), Lewis (1986) and Dai (1989), etc.

$$\bar{E}\dot{x}(t) = \bar{A}x(t) + \bar{B}u(t), \quad (1.1a)$$

$$y(t) = \bar{C}x(t), \quad (1.1b)$$

where $\bar{E}$, $\bar{A} \in \mathbb{R}^{n\times n}$, $\bar{B} \in \mathbb{R}^{n\times m}$ and $\bar{C} \in \mathbb{R}^{p\times n}$. The g.s.s system (1.1) for the special case where $\bar{E} = I$ is called regular and for the special case where $\det \bar{E} = 0$ is called singular. Systems of the form (1.1) cover a variety of cases in many areas of science and technology such as electrical circuits (Newcomb, 1982) network theory (Stott, 1979), power systems (Campbell and Rose, 1982), nuclear reactors (Reddy and Sannuti, 1975), robotics (Mills and Goldenberg, 1989), aircraft systems (Stevens, 1984), economical systems (Luenberger and Arbel, 1977), neurological events (Zeeman, 1976), catastrophic behavior (Sastry and Desoer, 1981), demography (Leslie model) (Campbell, 1980), etc.

Generic controllability, as well as all other problems which refer to generic system properties, refer to physical systems whose true values of their parameters are characterized by small deviations from their nominal values. Hence, the parameters of the system may be either fixed zeros or indeterminate (arbitrary) entries. Thus, the coefficients of the system matrices may be considered either fixed at zero or mutually independent free parameters. A matrix in such a form is called a structured matrix (Wonham, 1974; Shields and Pearson, 1976).

The form of a structured matrix is directly related to the physical topology of a system. In other words, even though there are many equivalent mathematical descriptions of a system, there must be one in which the elements of the system matrices are in structured form. Most of the rest of the mathematical descriptions may have system matrices in which the free parameters are bound among themselves.

For the case of g.s.s. systems the property of generic controllability has been studied only for discrete, casual systems. For this type of systems the following form is argued to be structured (Yamada and Luenberger, 1985)

$$Ex_{k+1} = Ax_k + Bu_k. \quad (1.2)$$

For the case of continuous g.s.s. systems no results have as yet been reported. This paper is devoted to establishing the first results on the subject and in particular the necessary and sufficient conditions for generic controllability of continuous g.s.s. systems. These results are analogous to those reported in Shields and Pearson (1976) for regular systems. Note that the causality property for the discrete system (1.2) facilitated the solution of the problem in Yamada and Luenberger (1985, 1986). For the continuous case, the analogous assumption can not be made. For this reason, the generic controllability problem for the continuous case is inherently more difficult to solve than that for the discrete case.

The proposed structured description for g.s.s. systems is of the following form

$$x(t) = \hat{E}\dot{x}(t) + \hat{A}x(t) + \hat{B}u(t), \quad y(t) = \hat{C}x(t). \quad (1.3)$$

Clearly, (1.3) is mathematicaly equivalent to (1.1). Note that description (1.3) is often used for describing many practical systems (large scale production (Luenberger and Arbel, 1977; Campbell, 1980)).

To study the generic controllability of g.s.s. systems, a duality property is proposed relating regular and g.s.s. systems. This duality property is based on a transformation which gives an alternate description of the g.s.s. systems. This is an isomorphism between regular and g.s.s. systems. This isomorphism makes it possible to take advantage of the knowledge we have for regular systems, to study the generic controllability of g.s.s. systems. Based on our approach for the study of the generic controllability of g.s.s. we readily establish the necessary and sufficient conditions for the generic observability of g.s.s. systems.

It is mentioned that the results of the present paper is part of the material reported in Koumboulis (1991).

2. *Duality property*

The g.s.s. system (1.1) in the frequency domain is

$$s\bar{E}X(s) = \bar{A}X(s) + \bar{E}x(0-) + \bar{B}U(s), \quad Y(s) = \bar{C}X(s). \quad (2.1)$$

The system is assumed solvable, i.e. it is assumed that $\det[s\bar{E} - \bar{A}] \not\equiv 0$. For this case, let μ be a real number such that $\det[\mu\bar{E} + \bar{A}] \neq 0$. Such a real number always exists and may take any value except the real roots $(\mu \neq \lambda)$ of $\det[(-\lambda)\bar{E} - \bar{A}] = 0$. The number of these roots is less than or equal to n. If we add to both sides of (2.1a) the quantity $\mu\bar{E}X(s)$ and after premultiplying the resulting relation by $(\mu\bar{E} + \bar{A})^{-1}$ then (2.1) becomes

$$(s+\mu)EX(s) = X(s) + Ex(0-) + BU(s), \quad Y(s) = CX(s), \quad (2.2)$$

where

$$E = (\mu\bar{E} + \bar{A})^{-1}\bar{E}, \quad B = (\mu\bar{E} + \bar{A})^{-1}\bar{B}, \quad C = \bar{C}.$$

* Received 31 May 1991; revised 28 June 1992; received in final form 23 August 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor V. Kučera under the direction of Editor H. Kwakernaak.

† National Technical University of Athens, Division of Computer Science, Department of Electrical Engineering, 15773 Zographou, Athens, Greece.

Next, consider the regular system

$$\dot{z}(t) = Ez(t) + Bv(t), \quad \psi(t) = Cz(t),$$

or equivalently in the Laplace domain w

$$wZ(w) = EZ(w) + z(0-) + BV(w), \quad \Psi(w) = CZ(w), \quad (2.3)$$

where $Z(w)$, $V(w)$ and $\Psi(w)$ are the Laplace transforms of $z(t)$, $v(t)$ and $\psi(t)$, respectively. The solution of system (2.2) for $X(s)$ and $Y(s)$ is

$$X(s) = [(s+\mu)E - I]^{-1}Ex(0-) + [(s+\mu)E - I]^{-1}BU(s), \quad (2.4a)$$

$$Y(s) = C[(s+\mu)E - I]^{-1}Ex(0-) + C[(s+\mu)E - I]^{-1}BU(s), \quad (2.4b)$$

and the solution of system (2.3) for $Z(w)$ and $\Psi(w)$ is

$$Z(w) = (wI - E)^{-1}z(0-) + (wI - E)^{-1}BV(w), \quad (2.5a)$$

$$\Psi(w) = C(wI - E)^{-1}z(0-) + C(wI - E)^{-1}BV(w). \quad (2.5b)$$

It is remarked that the two systems (2.2) and (2.3) involve three matrices, namely, the matrices E, B and C, which are identical for both systems. Due to this characteristic, these two systems are strongly interrelated. If we let

$$w = \frac{1}{s+\mu}, \quad s = \frac{1}{w} - \mu, \quad V(w) = U\left(\frac{1}{w} - \mu\right),$$
$$z(0-) = Ex(0-), \quad (2.6)$$

then, according to (2.4) and (2.5), we derive the following relations

$$X(s) = -wZ(w), \quad Z(w) = -(s+\mu)X(s), \quad (2.7a)$$

$$Y(s) = -w\Psi(w), \quad \Psi(w) = -(s+\mu)Y(s). \quad (2.7b)$$

Also, the transfer function matrices $H_g(s)$ and $H_r(s)$ of systems (2.2) and (2.3), respectively, are related as follows

$$H_g(s) = -wH_r(w), \quad H_r(w) = -(s+\mu)H_g(s). \quad (2.8)$$

From (2.7) and (2.8) it is observed that, for every g.s.s. system (2.1), which can always be written in the form (2.2), there exists a regular system of the form (2.3) having the characteristic in that it behaves similarly to the g.s.s. system (2.2) after a frequency reversion. These observations reveal that there exists a strong interrelationship between systems (2.2) and (2.3). This interrelationship is called here duality property and the two systems are called dual. It is noted that description (2.2) has the advantage over that reported in Zhou *et al.* (1987), Pandolfi (1980) and Christodoulou and Paraskevopoulos (1985) in that for its derivation no state vector transformations are required. Finally, it is remarked that description (1.3) is a special form of description (2.1) (let $\bar{E} = \hat{E}$, $\bar{A} = I - \hat{A}$, $\bar{B} = -\hat{B}$ and $\bar{C} = \hat{C}$).

Before closing this section, the equivalence between the controllability of the g.s.s. system (2.2) and its dual regular system (2.3) will be proven using the duality property presented above. This result will be useful for the study of the generic controllability that follows. To this end, we refer to the results reported in Cobb (1984), where it is shown that the g.s.s. system (2.1) is controllable if and only if

$$\operatorname{Im}(\lambda\bar{E} - \bar{A}) + \operatorname{Im}\bar{B} = \mathbb{R}^n, \quad \forall\lambda \in \mathbb{C} \quad \text{and} \quad \operatorname{Im}\bar{E} + \operatorname{Im}\bar{B} = \mathbb{R}^n.$$

The above conditions may obviously be written as

$$\operatorname{rank}[\lambda\bar{E} - \bar{A} \,\vdots\, \bar{B}] = n, \quad \forall\lambda \in \mathbb{C} \quad \text{and} \quad \operatorname{rank}[\bar{E} \,\vdots\, \bar{B}] = n. \quad (2.9)$$

If we let $s = \lambda - \mu$ and upon premultiplying the matrices in the rank operators in (2.9) by the matrix $(\mu\bar{E} + \bar{A})^{-1}$, conditions (2.9) may be expressed equivalently by the following two conditions

$$\operatorname{rank}[(s+\mu)E - I \,\vdots\, B] = n, \quad \forall s \in \mathbb{C}, \quad (2.10a)$$

$$\operatorname{rank}[E \,\vdots\, B] = n. \quad (2.10b)$$

Since, for $s = -\mu$, condition (2.10a) is an identity, it suffices to hold $\forall s \in \mathbb{C} - \{\mu\}$. The matrix $S_m(w)$ given by

$$S_m(w) = \begin{bmatrix} -wI_n & 0 \\ 0 & I_m \end{bmatrix}, \quad w = \frac{1}{s+\mu}, \quad (2.11)$$

is well defined for every $s \neq -\mu$. If we postmultiply the matrix in the rank operator in (2.10a) by $S_m(w)$, relation (2.10a) may be written as

$$\operatorname{rank}[wI - E \,\vdots\, B] = n, \quad \forall w \in \mathbb{C} - \{0\}, \quad w = \frac{1}{s+\mu}. \quad (2.12)$$

Combining relations (2.10b) and (2.12) we conclude that the criterion for the controllability of system (2.1) is

$$\operatorname{rank}[wI - E \,\vdots\, B] = n, \quad \forall w \in \mathbb{C}. \quad (2.13)$$

Relation (2.13) is also the controllability criterion for the dual regular system (2.3) (Kailath, 1980). We have thus proven that system (2.1) (or equivalently (2.2)) is controllable if and only if its dual regular system (2.3) is controllable, i.e. if and only if

$$\operatorname{rank}[B \,\vdots\, EB \,\vdots \cdots \vdots\, E^{n-1}B] = n. \quad (2.14)$$

Using the duality property we may also prove the equivalence between the observability (Cobb, 1984), of the g.s.s. system (2.1) and its dual regular system (2.3), i.e. the equivalence between the observability of the g.s.s. system (2.1) and the criterion

$$\operatorname{rank}[C^T \,\vdots\, E^TC^T \,\vdots \cdots \vdots\, (E^T)^{n-1}C^T] = n. \quad (2.15)$$

It is mentioned that criteria (2.14) and (2.15) have also been derived in Shayman and Zhou 1987) using a different approach.

3. *Generic controllability*

This section is devoted to deriving a generic criterion (Wonham, 1974) for the controllability of g.s.s. systems, under the assumption that system (1.3) is structured.

Definition 3.1. The g.s.s. system (1.3) is called structured if its system matrix $[\hat{E} \,\vdots\, \hat{A} \,\vdots\, \hat{B} \,\vdots\, \hat{C}]$ is a structured matrix. Note that a matrix is structured if its elements are either fixed at zero or mutually independent free parameters.

Let N be the number of the free parameters that appear in the matrix $[\hat{E} \,\vdots\, \hat{A} \,\vdots\, \hat{B} \,\vdots\, \hat{C}]$ and let ξ denote the N-dimensional vector of these parameters. Partition ξ as follows

$$\xi = [\underbrace{\phi}_{N_1} \,\vdots\, \underbrace{r}_{N_2} \,\vdots\, \underbrace{\pi}_{N_3} \,\vdots\, \underbrace{\sigma}_{N_4}], \quad N_1 + N_2 + N_3 + N_4 = N, \quad (3.1)$$

where ϕ, r, π and σ are the free independent vectors of the parameters appearing in $\hat{E}$, $\hat{A}$, $\hat{B}$ and $\hat{C}$, respectively, i.e. $\hat{E} = \hat{E}(\phi)$, $\hat{A} = \hat{A}(r)$, $\hat{B} = \hat{B}(\pi)$ and $\hat{C} = \hat{C}(\sigma)$.

In order to facilitate the generic manipulation of (1.3) we present three definitions, one proposition (Proposition 3.1) and prove one lemma (Lemma 3.1).

Definition 3.2 (Wonham, 1974). A variety $V \subset \mathbb{R}^n$ is defined to be the locus of common zeros of a finite number of polynomials $q_1, q_2, \ldots, q_k$:

$$V = \{\xi, q_i(\xi_1, \ldots, \xi_N) = 0, \ i = 1, 2, \ldots, k\}, \quad (3.2)$$

V is proper if $V \neq \mathbb{R}^N$.

Definition 3.3 (Shields and Pearson, 1976). We will say that a property Π, $(\Pi(\xi): \mathbb{R}^N \to \{0, 1\})$ is generic if there exist a proper variety V such that: If

$$\Pi(\xi) = 0, \quad \text{then} \quad \xi \in V. \quad (3.3)$$

Definition 3.4 (Yamada and Luenberger, 1985). An $m \times n$ matrix $M(\xi)$ is said to be a column structured matrix (C.S.M) if:

(i) each element of $M(\xi)$ is either fixed at zero, or a rational function of the free parameters $\xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^N$ and;

(ii) there exists a set of rational functions $q_i(\cdot,\cdot)$: $\mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$, $i = 1, \ldots, n$, such that, for $i = 1, 2, \ldots, n$,

$$M(\xi)\,\mathrm{diag}\,(1, \ldots, 1, a, 1, \ldots, 1) = M(q_i(\xi, a)) \quad \text{and} \quad q_i(\xi, 1) = \xi. \quad (3.4)$$

Proposition 3.1 (Yamada and Luenberger, 1985). A structured matrix is always a C.S.M.

Lemma 3.1 The matrix $I - \hat{A}$ is generically nonsingular.

Proof. Since the matrix $\hat{A}$ has been assumed structured, it is obvious according to Definition 3.2 that it is also a C.S.M. Hence, the matrix $\hat{A}$ has $\nu(\hat{A})$ generically nonzero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{\nu(\hat{A})}$ which are mutually distinct (Yamada and Luenberger, 1985). Here $\nu(\hat{A})$ denotes the generic order of $\hat{A}$, and it is defined by

$$\nu(\hat{A}) = \max\left\{k \;\middle|\; \sum_{i_1,\ldots,i_k} \det(\hat{A}[i_1, i_2, \ldots, i_k]) \neq 0\right\}, \quad (3.5)$$

where $\hat{A}[i_1, i_2, \ldots, i_k]$ is the submatrix of $\hat{A}$ consisting of the columns and rows $i_1, i_2, \ldots, i_k$ and the summations range over all possible combinations of integers satisfying $1 < i_1 < i_2 < \cdots < i_k$. Hence we have

$$\det[I - \hat{A}] = (1 - \lambda_1)(1 - \lambda_2)\cdots(1 - \lambda_\nu \quad (3.6)$$

Define

$$q(r) = \det[I - \hat{A}(r)].$$

Clearly, since $\hat{A}(r)$ has been assumed structured, the function $q(r)$ is a polynomial in the elements of r. Thus the values of r, for which $q(r) = 0$, constitute a variety V in $\mathbb{R}^{N_2}$. This variety is proper ($V \neq \mathbb{R}^{N_2}$), since for $r = 0$ we have that $\hat{A}(r) = 0$ and hence from (3.6) we have $q(0) = 1$. Define the property

$$\Pi(\hat{A}(r)) = \begin{cases} 0, & \text{if} \quad q(r) = 0 \\ 1, & \text{if} \quad q(r) \neq 0 \end{cases}$$

From this property it follows that $\Pi(\hat{A}(r)) = 0$, $\forall r \in V$ and $\Pi(\hat{A}(r)) = 1$, $\forall r \in V^c$. Thus $|I - \hat{A}(r)| \neq 0$ $\forall r$: $r \in V^c$. From this last result, together with Definition 3.3, we conclude that the matrix $I - \hat{A}(r)$ is generically nonsingular. ∎

Lemma 3.1 is important since it allows the manipulation of system (1.3) in the generic sense. Indeed, using Lemma 3.1, system (1.3) is expressed equivalently as

$$E^*\dot{x} = x + B^*u, \quad y = \hat{C}x, \quad E^* = D^{-1}\hat{E},$$
$$B^* = -D^{-1}\hat{B}, \quad D = I - \hat{A}. \quad (3.7)$$

In the sequel, instead of studying the original matrices $\hat{E}$, $\hat{A}$ and $\hat{B}$ of system (1.3), we will be studying the matrices E^* and B^* of system (3.7). The matrices E^* and B^* have the property that they are C.S.M., a fact which will be useful in proving the necessary and sufficient conditions for generic controllability. This property is proven in the following lemma.

Lemma 3.2. The matrices E^* and B^* are C.S.M.

Proof. According to definition (3.1) and relation (3.7) we may write

$$E^* = E^*([\phi, r]) = D^{-1}(r)\hat{E}(\phi). \quad (3.8)$$

From the definition $D = I - \hat{A}$, we readily observe that the elements of the matrix $D(r)$ are structural zeros or polynomials of r. Thus, the elements of $D^{-1}(r)$ are structural zeros or ratios having in their numerators and denominators polynomials of r. Using this last observation, together with the assumption that the matrix $\hat{E}(\phi)$ is a structured matrix (Definition 3.1), we readily conclude that the elements of the matrix $E^*([\phi, r]) = D^{-1}(r)\hat{E}(\phi)$ are structural zeros or sums of products involving rational functions of elements of r and ϕ. Thus, the elements of E^* are structural zeros or rational functions of $[\phi, r]$. Hence, the first of the conditions, appearing in Definition 3.4, is satisfied.

In order to investigate if the matrix E^* satisfies the second condition of Definition 3.4, we start by noting that since $\hat{E}(\phi)$ is a structured matrix, then according to Proposition 3.1 it is also a C.S.M. Hence, there exists a set of rational functions $\hat{q}_i(\phi, a)(\hat{q}_i(\cdot,\cdot)$: $\mathbb{R}^{N_1} \times \mathbb{R} \rightarrow \mathbb{R}^{N_1})$, $i = 1, \ldots, n$ such that

$$\hat{E}(\phi)\,\mathrm{diag}\,(1, \ldots, 1, a, 1, \ldots, 1) = \hat{E}(\hat{q}_i(\phi, a)), \quad (3.9a)$$

and

$$\hat{q}_i(\phi, 1) = \phi. \quad (3.9b)$$

Defining the rational functions

$$q_i^*([\phi, r], a): \mathbb{R}^{N_1 + N_2} \times \mathbb{R} \rightarrow \mathbb{R}^{N_1 + N_2}$$
$$q_i^*([\phi, r], a) = [\hat{q}_i(\phi, a), r], \quad (3.10)$$

for $i = 1, \ldots, n$ and using (3.8)–(3.10), one may derive that

$$\begin{aligned} E^*([\phi, r])\,\mathrm{diag}\,(1, \ldots, 1, a, 1, \ldots, 1) &= D^{-1}(r)\hat{E}(\phi)\,\mathrm{diag}\,(1, \ldots, 1, a, 1, \ldots, 1) \\ &= D^{-1}(r)\hat{E}(\hat{q}_i(\phi, a)) \\ &= E^*([\hat{q}_i(\phi, a), r]) \\ &= E^*(q_i^*([\phi, r], a)), \end{aligned} \quad (3.11)$$

where a is in the ith position. Finally, it is remarked that from the definitions of q_i^* appearing in (3.10), and using (3.9b), we have that $q_i^*([\phi, r], 1) = [\hat{q}_i(\phi, 1), r] = [\phi, r]$. Thus, Lemma 3.2 has been proven with regard to E^*. The proof for B^* is directly analogous to that for E^*. ∎

Further, we present the following two definitions (Shields and Pearson, 1976).

Definition 3.5. The g-rank($\cdot$) denotes the generic rank of a matrix and we say that the generic rank of $[\hat{E} \mid \hat{B}]$ equals to n if the property

$$\Pi([\hat{E}(\phi) \mid \hat{B}(\pi)]) = \begin{cases} 0, & \text{if rank}\,[\hat{E}(\phi) \mid \hat{B}(\pi)] < n, \\ 1, & \text{if rank}\,[\hat{E}(\phi) \mid \hat{B}(\pi)] = n, \end{cases} \quad (3.12)$$

is generic.

Definition 3.6. An $n \times (n + m)$ matrix $[E^* \mid B^*]$ is reducible if for some permutation matrix P the following holds

$$[P^TE^*P \mid P^TB^*] = \begin{array}{c} \begin{array}{ccc} \overbrace{\quad}^{p} & \overbrace{\quad}^{n-p} & \overbrace{\quad}^{m} \end{array} \\ \left[\begin{array}{c|c|c} * & 0 & 0 \\ \hline * & * & * \end{array}\right] \end{array}, \quad 1 \le p \le n. \quad (3.13)$$

It is irreducible if it is not reducible.

Using all above results, we next establish the following theorem which constitutes the main contribution of the present paper.

Theorem 3.1. The structured g.s.s. system (1.3) is generically controllable if and only if

(i) g-rank $[\hat{E} \mid \hat{B}] = n$(full), (3.14a)

(ii) $[D^{-1}\hat{E} \mid D^{-1}\hat{B}]$ is irreducible. (3.14b)

Proof. As it has been indicated in Section 2 a g.s.s. system is controllable if and only if its dual regular system is controllable. Thus, system (3.7) (or equivalently system (1.3)) is controllable if and only if its dual regular system

$$\dot{z}(t) = E^*z(t) + B^*v(t), \quad (3.15)$$

is controllable. Therefore, investigating the generic controllability of system (1.3) is equivalent to investigating the generic rank of the matrix $[B^* \mid E^*B^* \mid \ldots \mid (E^*)^{n-1}B^*]$. To investigate the generic controllability of (3.15), we start by noting that in Lemma 3.2 it was proven that the matrices E^* and B^* of the regular system (3.15) are C.S.M. Applying the results in Yamada and Luenberger (1985) for a regular system with C.S.M. system matrices to system (3.15), we have that the necessary and sufficient conditions for the

generic controllability of (3.15) are

$$g\text{-rank}\,[E^* \mid B^*] = n(\text{full}), \tag{3.16a}$$

$$[E^* \mid B^*] \text{ is irreducible.} \tag{3.16b}$$

Substituting the definitions E^* and B^* appearing in (3.7) we observe that (3.16b) reduces to (3.14b), whereas (3.16a) takes on the form

$$g\text{-rank}\,[D^{-1}\hat{E} \mid D^{-1}\hat{B}] = n, \tag{3.17}$$

or equivalently

$$g\text{-rank}\,\{D^{-1}[\hat{E} \mid \hat{B}]\} = n. \tag{3.18}$$

Clearly, since D is generically nonsingular (Lemma 3.1) and its parameters $r_1, \ldots, r_{N_2}$ are all different than $\phi_1, \ldots, \phi_{N_1}$ and $\pi_1, \ldots, \pi_{N_3}$ of the structured matrices $\hat{E}$ and $\hat{B}$, Condition (3.18) reduces to that in (3.14a). Thus, Theorem 3.1 has been proven. ∎

Before closing this section, we focus our attention to the solvability of system (1.3). From Lemma 3.1 we have that the matrix $I - \hat{A}$ is generically nonsingular. This means that the characteristic polynomial $\det[s\hat{E} + \hat{A} - I]$ is not identically equal to zero since, for $s = 0$, it reduces to $\det[I - \hat{A}]$. Thus description (1.3) with $\hat{E}$, $\hat{A}$ and $\hat{B}$ structured matrices is always generically solvable.

4. *Generic observability*

Based on our results for the generic controllability of g.s.s. systems we will now establish the necessary and sufficient conditions for the generic observability of g.s.s. systems. Description (1.3) is generically observable if and only if description (3.7) is. As it has been noted in Section 2, a g.s.s. system is observable if and only if its dual regular system is also observable. Therefore investigation of the generic observability of system (1.3) is equivalent to that of investigating the generic rank of the matrix $[\hat{C}^T \mid (E^*)^T\hat{C}^T \mid \ldots \mid [(E^*)^T]^{n-1}\hat{C}^T]$. Following the steps of the proof of Theorem 3.1 we readily conclude that g-rank $[\hat{C}^T \mid (E^*)^T\hat{C}^T \mid \ldots \mid [(E^*)^T]^{n-1}\hat{C}^T] = n$ if and only if

$$g\text{-rank}\,[(E^*)^T \mid \hat{C}^T] = n(\text{full}) \tag{4.1}$$

$$[(E^*)^T \mid \hat{C}^T] \text{ is irreducible.} \tag{4.2}$$

Since D is generically nonsingular, then substitution of (3.8) in (4.1 and 2) yields

$$g\text{-rank}\begin{bmatrix}\hat{E} \\ \hline \hat{C}\end{bmatrix} = n(\text{full}), \tag{4.3}$$

$$\begin{bmatrix}D^{-1}\hat{E} \\ \hline \hat{C}\end{bmatrix}^T \text{ is irreducible.} \tag{4.4}$$

We have thus established the following corollary.

Corollary 4.1. The structured g.s.s. system (1.3) is generically observable if and only if Conditions (4.3) and (4.4) are satisfied.

5. *Conclusions*

In this paper the necessary and sufficient conditions for the generic controllability and observability of continuous g.s.s. systems are established (Theorem 3.1 and Corollary 4.1, respectively). The proposed approach appears to be a very useful tool for studying other related problems such as the generic stabilizability, detectability and stability of g.s.s. systems, as well as the generic controllability and observability of implicit systems. These problems are currently under investigation.

References

Campbell, S. L. (1980). *Singular Systems of Differential Equations*. Pitman, San Francisco.

Campbell, S. L. and N. J. Rose (1982). A second-order singular linear system arising in electric power systems analysis. *Int. J. System Science*, **13**, 101–108.

Christodoulou M. A. and P. N. Paraskevopoulos (1985). Solvability, controllability and observability of singular systems. *J. Optimiz. Theory and Applications*, **45**, 53–57.

Cobb, J. D. (1984). Controllability, observability and duality in singular systems. *IEEE Trans. Aut. Control*, **AC-29**, 1076–1082.

Dai, L. (1989). *Singular Control Systems*. Springer–Verlag, Berlin.

Kailath, T. (1980). *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Koumboulis, F. N. (1991). New techniques for analysis and feedback design for regular and singular systems, Ph.D. Thesis, National Technical University of Athens, Department of Electrical Engineering, Greece.

Lewis, F. L. (1986). A survey of linear singular systems. *Circuits, Systems and Signal Processing*, **5**, 3–36.

Luenberger, D. G. and A. Arbel (1977). Singular dynamic Leontieff systems. *Econometrica*, **45**, 991–995.

Mills, J K. and A. A. Goldenberg (1989). Force and position control of manipulators during constrained motion tasks. *IEEE Trans. Robotics and Automation*, **RA-5**, 3–46.

Newcomb, R. W. (1982). Semistate design theory: binary and swept hysteresis. *Circuits, Systems and Signal Processing*, **1**, 205–216.

Pandolfi, L. (1980) Controllability and stabilizability for linear systems of algebraic and differential equations. *J. Optimization Theory and Applications*, **30**, 601–620.

Reddy, P. B. and P. Sannuti (1975). Optimal control of a coupled core nuclear reactor by a singular perturbation method *IEEE Trans. Aut. Control*, **AC-20**, 766–769.

Sastry, S. S. and C. A. Desoer (1981). Jump behavior of circuits and systems. *IEEE Trans. Circuit and Systems*, **CAS-28**, 1109–1123

Shayman, M. A. and Z. Zhou (1987). Feedback control and classification of generalized linear systems. *IEEE Trans. Aut. Control*, **AC-32**, 483–494.

Shields, R. W. and J. B. Pearson (1976). Structural controllability of multi-input multi-output linear systems. *IEEE Trans. Aut. Control*, **AC-21**, 203–212.

Stevens, B. L. (1984). Modeling, simulation, and analysis with state variables. Report LG84RR002, Lockheed-Georgia Co., Marietta, GA.

Stott, B. (1979). Power system response dynamic calculations. *Proc. IEEE*, **67**, 219–241.

Verghese, G. C., B. C. Levy and T. Kailath (1981). A generalized state-space for singular systems. *IEEE Trans. Aut. Control*, **AC-27**, 811–831.

Wonham, W. M. (1974). *Linear Multivariable Control: A Geometric Approach*. Springer-Verlag, Berlin.

Yamada, T. and D. G. Luenberger (1985). Generic controllability theorems for descriptor systems. *IEEE Trans. Aut. Control*, **AC-30**, 144–152

Yamada, T. and D. G. Luenberger (1985). Algorithms to verify generic causality and controllability of descriptor systems. *IEEE Trans. Aut. Control*, **AC-30**, 874–880.

Zeeman, E. C. (1976). Duffin's equation in brain modeling. *J. Inst. Math. and App.*, **28**, 207–214.

Zhou, Z., M. A. Shayman and J.-J. Tarn (1987). Singular systems; a new approach in the time-domain. *IEEE Trans. Aut. Control*, **AC-32**, 42–50.

Automatica, Vol. 29, No. 2, pp. 531–536, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Brief Paper

Reachable, Controllable Sets and Stabilizing Control of Constrained Linear Systems*

J. B. LASSERRE†

Key Words—Discrete-time systems; linear systems; controllability; constraint theory.

Abstract—We consider discrete-time linear systems with constrained controls. We present an algorithm for computing an external representation of the N-step reachable set and controllable set. It is an improvement of previous methods. In particular, it is shown that for constructing the set of initial states that can be driven to the origin asymptotically, one can restrict the search to a polyhedral set in the unstable subspace of the autonomous system, which yields a drastic reduction in the computational burden. Application to the synthesis of stabilizing feedback controllers is discussed.

1. *Introduction*

IN THIS PAPER, we are concerned with discrete-time linear systems with *constrained controls*, $x_{k+1} = Ax_k + Bu_k$, $\underline{u} \le u_k \le \bar{u}$. Polyhedral constraints on the state can also be incorporated. Such discrete-time models can be inherently discrete systems or sampled-data systems or even approximations of continuous systems. Most real life systems have physical limitations on the controls and surprisingly enough, apart from papers related to optimal control issues, few studies directly incorporate this feature. For example, the design of feedback controllers often ignores this feature and when a control saturates its bound, the linear system exhibits a nonlinear behavior. Among possible consequences, stability properties are difficult to analyse and linear properties (for state estimation for instance) are lost.

In the design of feedback controllers that take into account control constraints, some polyhedral set (necessarily included in the controllable set) must be positively invariant (see for instance Benzaouia and Burgat (1988), Bitsoris (1988), Hennet and Beziat (1991), Hennet and Castelan (1991), Pajunen and Erdol (1990), Vassilaki and Bitsoris (1989) and Vassilaki *et al.* (1988)). Choosing such a set is not trivial (Blanchini, 1990b). Ideally, one would try to build a controller such that the controllable set is positively invariant. Therefore, comparison of this invariant set and the controllable set may yield a measure of the quality of this controller (the larger this set the larger the *number* of states that can be driven to the origin). Also, knowledge of the reachable or controllable set helps to solve minimum time problems.

Concerning reachable and controllable sets, previous results in the literature provide numerical procedures to test whether some given state is controllable or not (see Barmish and Schmitendorf (1980) and Van Til and Schmitendorf (1986) for example) (and sometimes in a more general framework) while others try to explicitly characterize the set of controllable states (see for example Desoer and Wing (1961), Fisher and Gayek (1990), Gutman and Cwikel (1987), Hamza and Rasmy (1971), Lasserre (1987), Lasserre (1991a), Rumchev and James (1989) and Witsenhausen (1972)). Our contribution falls into the latter category of papers.

We first determine an exact external representation (i.e. bounding hyperplanes) of the N-step reachable set from any given initial state and the controllable set, i.e. the set of initial states from which the system can be driven to zero in N-steps. It is worth noting that if zero is an admissible control, the N-step controllable set is an inner approximation (the larger N the better the approximation) of the set of all controllable states, which is not polyhedral in general. Our method is different from others in the literature (Hamza and Rasmy (1971); Lin (1970); Gutman and Cwikel (1987)) and is less demanding in terms of computational burden, a critical issue in this case. An interesting exception concerns the two-dimensional case where an exact characterization with a closed-form expression has been provided with an algorithm of polynomial complexity. (See Lasserre (1991a) for a representation in terms of bounding hyperplanes (external representation) and Fisher and Gayek (1990) in terms of vertices (internal representation).)

More importantly, after some analysis, we show that it suffices to restrict attention to the unstable subspace of the autonomous system. Once a controllable set C is determined in this subspace, it is shown that all the states in a cylinder (whose projection in the unstable subspace is C) can be driven to the origin asymptotically. This technique can yield substantial computational savings since the dimension of the problem is decreased. For example, if there are only two unstable eigenvalues, the problem reduces to finding a controllable set in dimension two for which a closed-form expression and a computational procedure in polynomial time exist (see again Lasserre (1991)).

Finally, in the last section we present a technique to build a feedback controller that satisfies the control constraints. Given an initial state whose projection in the unstable subspace is in C, we first drive this projection to a smaller set C' by using a simple feedback control. Then, using simple linear algebra techniques as in Hennet and Beziat (1991), a stabilizing linear state-feedback controller is built such that a set (whose projection in the unstable subspace is the set C') is positively invariant for the closed-loop system.

2. *Preliminary results*

In this section we present preliminary results that will be extensively used in the computational procedure of the N-step reachable and controllable sets, presented in the next section. We show how to compute in a simple manner the Minkowski sum of a convex polytope and a line segment in R^n. The bounding hyperplanes of the resulting polytope are deduced from the original polytope by simple formula.

Consider an (m, n) matrix A and a vector $b \in R^m$ such that $\{x \in R^n \mid Ax \le b\}$ defines a nonempty (bounded) convex polytope Ω. Let h be some vector in R^n. Let us recall that the Minkowski sum $\Omega_1 + \Omega_2$ of two convex polytopes Ω_1 and Ω_2 is the convex polytope $\{x \mid \exists y \in \Omega_1,\ z \in \Omega_2,\ x = y + z\}$. Let $I(\Omega)$ be the set of integers $\{1, 2, \ldots, r\}$, which are the row indices of the matrix A. Let y^i be the vector with ith

* Received 6 June 1991; revised 2 January 1992; received in final form 9 March 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor T. Başar under the direction of Editor H. Kwakernaak.

† LAAS-CNRS, 7 Avenue du Colonel Roche, 31 077 Toulouse, France.

component $\max[0, y_i]$. Moreover let vertex (Ω) be the set defined as:

$$i_1, \ldots, i_p \in I(\Omega); \quad (i_1, i_2, \ldots, i_p) \in \text{vertex}(\Omega)$$
$$\Leftrightarrow \exists x \in \Omega, \quad A_{i_k}x = b_{i_k}, \quad \forall k = 1, \ldots, p,$$

where A_{i_k} stands for the i_k row of matrix A and b_{i_k} for the i_k component of b.

Proposition 2.1. The Minkowski sum $S(\Omega, h)$ of the convex polytope Ω and the convex polytope $\{y \mid y = \lambda h, \lambda \in (0, 1)\}$, is the convex polytope defined by

$$Ax \le b + (Ah)^+, \tag{1}$$

$$(v_iA_i + v_jA_j)x \le v_ib_i + v_jb_j, \quad \forall (i, j) \in P(\Omega), \tag{2}$$

where $P(\Omega) \subset \text{vertex}(\Omega)$ is defined as

$$P(\Omega) = \{(i, j) \in \text{vertex}(\Omega) \mid (A_ih) . (A_jh) < 0,$$
$$\text{i.e. } \exists v_i, v_j \ge 0, (v_iA_i + v_jA_j)h = 0\}. \tag{3}$$

Proof. First we prove that $S(\Omega, h)$ is the polytope defined by (1) and (2) where in (2), instead of $P(\Omega)$ we consider the set $M(\Omega)$ of *all* the pairs (i, j) such that $(A_ih) . (A_jh) < 0$. Then we show that only those pairs in vertex (Ω) are worth considering.

Consider an x which satisfies $Ax \le b$. Then $A(x + \lambda h) \le b + (Ah)^+$ and $(v_iA_i + v_jA_j)x \le v_ib_i + v_jb_j$ for all (i, j) which satisfy $(A_ih) . A_jh) < 0$.

Conversely, consider a vertex x^* of the polytope defined by (8) and (9). It must satisfy:

$$A_ix^* = b_i + (A_ih)^+, \quad \forall i \in I(x^*) \subset I(\Omega),$$

$$(v_kA_k + v_pA_p)x = v_kb_k + v_pb_p, \quad \forall (k, p) \in M'(\Omega) \subset M(\Omega),$$

for some subsets $I(x^*)$, $M'(\Omega)$. Actually, either $A_ih \ge 0$ or $A_ih \le 0$ holds for all $i \in I(x^*)$ otherwise if $A_ih > 0$ and $A_jh < 0$ for some i, $j \in I(x^*)$ the constraint $v_iA_ix + v_jA_jx \le v_ib_i + v_jb_j$ is violated for the scalars v_i, v_j which satisfy $v_iA_ih + v_jA_jh = 0$.
(i) Consider the case where $A_ih > 0$, $\forall i \in I(x^*)$.
Then, $A_i(x^* - h) = b_i$, $\forall i \in I(x^*)$. Moreover, $A_k(x^* - h) \le b_k$ for all $k \notin I(x^*)$, $A_kh \ge 0$. For those k such that $A_kh < 0$ we cannot have $A_k(x^* - h) > b_k$ since otherwise the constraint $v_iA_ix + v_kA_k \le v_ib_i + v_kb_k$ is violated at x^* for the scalars v_i, v_k which satisfy $v_iA_ih + v_kA_kh = 0$, $i \in I(x^*)$. Hence $A(x^* - h) \le b$.
(ii) Consider now the case where $A_ih \le 0$ for all $i \in I(x^*)$.
$A_ix^* \le b_i$, $\forall i \in I(x^*)$ and for all $k \notin I(x^*)$, s.t. $A_kh \le 0$. Moreover, $A_ix^* > b_i$ for some i such that $A_ih > 0$ is impossible, since otherwise the constraint $(v_iA_i + v_kA_k)x \le v_ib_i + v_kb_k$ for which the scalars v_i, v_k satisfy $v_iA_ih + v_kA_kh = 0$, $k \in I(x^*)$ is violated at x^*. Hence $Ax^* \le b$.

In summary, any vertex x of $S(\Omega, h)$ is the sum of an element y of Ω and λh with λ being either 0 or 1. Therefore $S(\Omega, h)$ is the Minkowski sum $\Omega + \lambda h$. However, many constraints of the type (2) are in fact redundant.

As before, consider a vertex x^* of $\Omega + \lambda h$.
(i) If $A_ix^* = b_i + A_ih$, $A_ih \ge 0$, $\forall i \in I(x^*)$ then we have seen that $x^* - h$ is a vertex of Ω. A constraint of the type (2) is not strictly redundant if and only if it is binding at some vertex of $S(\Omega, h)$. Consider a binding constraint $(v_kA_k + v_pA_p)x$ at x^*. Since $x^* - h$ is a vertex of Ω, the constraint is also binding at $x^* - h$. But since $x^* - h$ is a vertex of Ω, this is possible if and only if $A_k(x^* - h) = b_k$ and $A_p(x^* - h) = b_p$ which means that $(k, p) \in \text{vertex}(\Omega)$ and therefore $(k, p) \in P(\Omega)$.
(ii) If $A_ix^* = b_i$, $\forall i \in I(x^*)$ (and $A_ih < 0$), then we know that x^* is also a vertex of Ω. Therefore, the only binding constraints $(v_iA_i + v_kA_k)x$ binding at x^* are those for which necessarily $A_ix^* = b_i$ and $A_kx^* = b_k$. But again this means that $(i, k) \in \text{vertex}(\Omega)$ and therefore $(i, k) \in P(\Omega)$, so that the proof is complete. □

An immediate consequence is the following:

Proposition 2.2. The Minkowski sum $S(\Omega, |h|)$ of the convex polytope Ω and the convex polytope $\{y \mid y = \lambda h, \lambda \in (-1, +1)\}$ is the convex polytope defined by

$$Ax \le b + |Ah|, \tag{4}$$

$$(v_iA_i + v_jA_j)x \le v_ib_i + v_jb_j, \quad \forall (i, j) \in P(\Omega). \tag{5}$$

Remark 2.1. Let us call A_h the matrix of the constraints which define $S(\Omega, |h|)$. Let $n(i, j)$ be the index of the constraint $(v_iA_i + v_jA_j)x \le v_ib_i + v_kb_k$. From the demonstration of Proposition 2.1 we also conclude that

- $(i, j) \in \text{vertex}(S(\Omega, |h|)) \Leftrightarrow (i, j) \in \text{vertex}(\Omega)$ and $(i, j) \notin P(\Omega)$
- $(i, n(i, j)) \in \text{vertex}(S(\Omega, |h|))$ and $(j, n(i, j)) \in \text{vertex}(S(\Omega, |h|))$
- $(i, n(p, q)) \in \text{vertex}(S(\Omega\, |h|)) \Leftrightarrow ((i, p)$ or $(i, q))$ and $(p, q) \in P(\Omega)$, and $(i, p, q) \in \text{vertex}(\Omega)$
- $(n(i, j), n(i, p)) \in \text{vertex}(S(\Omega, |h|)) \Leftrightarrow (i, j), (i, p) \in P(\Omega)$ and $(i, j, p) \in \text{vertex}(\Omega)$
- $(n(i, j), n(p, q) \in \text{vertex}(S(\Omega, |h|)) \Leftrightarrow (i, j), (p, q) \in P(\Omega)$, (i, p) and $(j, q) \in P(\Omega)$ or (i, q) and $(j, p) \in P(\Omega)$, and $(i, j, p, q) \in \text{vertex}(\Omega)$.

Therefore, to determine all the pairs of rows of A_h which are in vertex$(S(\Omega, |h|))$ it is not necessary to check all pairs. This property will be used to save computations in the procedure below.

3. *Reachable and controllable sets*

In this section we present a computational procedure to determine the boundary hyperplanes of the N-step reachable and controllable sets. No controllability assumption is required, and the matrix A is not required to be non-singular. If the system is not controllable, then the reachable and controllable sets will be degenerate, i.e. they will be contained in an affine variety of dimension strictly less than n. We will just observe that some constraints are of the form $gx \le c$ and $-gx \le -c$, i.e. they define the hyperplane $gx = c$. We also compare our procedure with others in the literature. For simplicity of notation we restrict to discrete time-invariant linear systems with bounded controls, but in the sequel, it will become obvious to the reader that the same methodology applies to non-stationary linear systems.

Let us consider a discrete time-invariant linear system with bounded controls

$$x_{k+1} = Ax_k + Bu_k, \quad \underline{u} \le u_k \le \bar{u}, \quad k = 1, \ldots, N, \tag{6}$$

where x is the state vector ($\in R^n$), u is the control vector ($\in R^m$), A is the matrix of the dynamics of the uncontrolled system, and $\underline{u}$, $\bar{u}$ are the (vectors) bounds on the control variables.

For any matrix M, M_i will stand for its ith row vector and for any vector v, $|v|$ will stand for the vector of absolute values.

3.1. *A computational procedure.*

Let us recall that the N-step reachable set is the set of $y \in R^n$ which satisfy

$$y - A^Nx_0 = \sum_{k=0}^{N-1} A^kBu_k, \quad |u_k| \le \bar{u}, \quad \forall k,$$

for some control sequence $u_0, \ldots, u_{N-1}$, or equivalently

$$y - A^Nx_0 = \sum_{k=0}^{\cdot} \sum_{p=1}^{\cdot\cdot} A^kB^{(p)}u_{kp}, \; u_{kp} \in (-\bar{u}_p, \bar{u}_p), \tag{7}$$

where $B^{(p)}$ stands for the pth column of B and where, to simplify notation, the index of control at period k has been switched from k to $N - (k + 1)$.

The last equation simply means that $y - A^Nx_0$ is the sum of Nm polytopes (line segments), each one of them being defined as λv where v is some vector of R^n and λ a scalar in some interval $(-a, +a)$. These polytopes are in fact *zonoids* (Bolker (1969); Witsenhausen (1972)). Therefore, since the Minkowski sum of polytopes is associative, the idea is to use Proposition 2.2 iteratively to compute the bounding hyperplanes of the reachable set.

Let $R(k, p)k \le N - 1$, $p \le m$, be the convex polytope

defined by

$$R(1,p)=\left\{x\in R^n \mid x=\sum_{j=1}^{p} B^{(j)}u_{0j},\quad u_{0j}\in-(\bar{u}_j,\bar{u}_j),\ \forall j\right\},$$

$$R(k,p)=\left\{x\in R^n \mid x=\sum_{i=0}^{k-2}\sum_{j=1}^{m} A^iB^{(j)}u_{ij}\right.$$

$$\left.+\sum_{l=1}^{p} A^{k-1}B^{(l)}u_{k-1l},\ u_{i,j}\in(-\bar{u}_j,\bar{u}_j),\ \forall i,j\right\},\ k\geq 2.$$

Therefore assume that

$$R(k,p)=\{x\in R^n \mid \Omega(k,p)x\leq b(k,p)\},$$

where $\Omega(k,p)$ is some $(r(k,p),n)$ matrix and $b(k,p)$ some $r(k,p)$ vector.

Then, according to Proposition 2.2, if $p<m$,

$$R(k,p+1)=S(R(k,p),|A^{k-1}B^{(p+1)}|\,\bar{u}_{p+1}),$$

and if $p=m$,

$$R(k+1,1)=S(R(k,m),|A^kB^{(1)}|\,\bar{u}_1),$$

where $S(.,.)$ is the Minkowski sum defined in Proposition 2.2.

One way to initialize the procedure is to compute $R(1,1)$ which is the convex polyhedron

$$\{x\in R^n \mid |x^TB^{(1)}|\leq \|B \qquad 0\leq u_i^Tx\leq 0,\ i=1,\ldots,n-1\},$$

where u_i, $i=1,\ldots,n-1$ are $n-1$ linearly independent vectors orthogonal to $B^{(1)}$, i.e. which satisfy $u_i^TB^{(1)}=0$, $i=1,\ldots,n-1$. Therefore we have:

Theorem 3.1. If $p>1$, $R(k,p)$ is the convex polytope (deduced from $R(k,p-1)$) defined as the set of all $x\in R^n$ such that

$$\Omega(k,p-1)x\leq b(k,p-1)+|\Omega(k,p-1)A^{k-1}B^{(p)}|\,\bar{u}_p, \tag{8}$$

$$(v_i\Omega_i(k,p-1)+v_j\Omega_j(k,p-1))x\leq v_ib_i(k,p-1)$$
$$+v_jb_j(k,p-1),\quad (i,j)\in P(R(k,p-1)), \tag{9}$$

and similarly if $p=1$, $R(k+1,1)$ is defined as the set of all $x\in R^n$ such that

$$\Omega(k,m)x\leq b(k,m)+|\Omega(k,m)A^kB^{(1)}|\,\bar{u}_1, \tag{10}$$

$$(v_i\Omega_i(k,m)+v_j\Omega_j(k,m))x\leq v_ib_i(k,m)$$
$$+v_jb_j(k,m),\quad (i,j)\in P(R(k,m)). \tag{11}$$

At each step, the number of bounding hyperplanes increases. This number depends on the cardinality of the set $P(R(k,p))$ which in turn highly depends on the dimension n of the state space. It is also immediate to check by induction that the right-hand-side $b(k,m)$ is always positive for all k, m. Moreover, since all the constraints on the controls are symmetrical, $R(k,p)$ is a symmetrical convex polytope for all k, p, i.e. the constraints are of the form $\pm cx\leq a$. Therefore, we immediately have the following.

Lemma 3.1. The N-step reachable set from initial state x_0 is the set of vectors $y\in R^n$ which satisfy:

$$|\Omega(N,m)(y-A^Nx_0)|\leq b(N,m), \tag{12}$$

and the N-step controllable set is the set C_N of vectors $y\in R^n$ which satisfy:

$$|\Omega(N,m)A^Ny|\leq b(N,m), \tag{13}$$

where $\Omega(N,m)$ has twice as few rows as in Theorem 3.1. Note that the N-step controllable step is an inner approximation of the set of all controllable states $\bigcup_{N\geq 1} C_N$ since if $x\in C_N$ then $x\in C_p$ for all $p\geq N$.

Remark 3.1. In the computational procedure, to obtain the hyperplanes in (9), a large number of pairs (i,j) need not be considered. It suffices to apply the rules described in Remark 2.1 to characterize vertex $(R(k,p))$, i.e. the pairs (i,j) of interest. Following these rules yields great computational savings.

Remark 3.2. Note that C_N is the set of all the states that can be driven to the origin in *finitely* many steps. When the system is not controllable, some states not included in C_N can be driven to the origin asymptotically (see Gutman and Cwikel (1987) for a simple illustrating example).

3.2. *Comparison with other procedures.* For linear systems with general *linear* constraints, a complete characterization of both reachable and controllable sets is presented in Lasserre (1987). The associated computational procedure, based on Farkas' lemma (Schrijver (1986)), is not realistic even for moderate size problems. For two-dimensional systems, computation is tractable and easy and even a closed-form expression of both sets is given in Lasserre (1991a). In the case treated here, Hamza and Rasmy (1971) have proposed a procedure to compute the bounding hyperplanes of the reachable set in the single input case with obvious changes for the multi-input case. Their procedure is an improvement of Lin's procedure in Lin (1970), where computing the extra hyperplanes of the type (9), (11) requires solving $\binom{mN}{n-2}$ (a large number even for reasonable N) linear systems of $n-1$ equations in $n-1$ unknowns. In Hamza and Rasmy (1971) still $\binom{mN}{n-2}$ linear systems need be considered but after a simple test, the new hyperplane eventually built is also a linear combination of two vectors like in our procedure. However, in our procedure: we need not consider those $\binom{mN}{n-2}$ linear systems. The hyperplanes are obtained directly from those of $R(k,p-1)$. The pairs of rows candidate to build a new hyperplane, are obtained by using Remark 2.3 without calculations.

In a more general setting, Gutman and Cwikel (1987) have proposed a procedure to compute controllable sets and reachables sets. They consider the case where the state is constrained to belong to some prescribed set X and the control is constrained to belong to some set Ω. However, their procedure requires knowledge of the vertices of Ω and at each step the set of vertices of $R(k,p)$. The vertices of $R(k+1,m)$ are deduced from those of $R(k,m)$ by adding in all possible ways a vertex of $R(k,m)$ and a vertex of Ω and then taking the convex hull of all these points, a tedious procedure. Moreover, the intersection with the set X must be computed and finally the vertices of this new polytope must also be computed which requires a highly combinatorial procedure with some linear algebra.

In the case where the control set is defined as $\{u\in R^m \mid \underline{u}\leq\Omega u\leq\bar{u}\}$ where Ω is a nonsingular square matrix, then by an obvious change of variables, one is back to the case $\underline{u}\leq u\leq\bar{u}$. Also, our procedure can incorporate state constraints of the type $Mx_k\leq a$. Indeed, once $R(k,m)$ has been computed it suffices to intersect it with $\{x \mid Mx\leq a\}$, i.e. the new $R(k,m)$ set is now defined through the inequalities:

$$\Omega(k,m-1)x\leq b(k,m-1)+|A^{k-1}B^{(m)}|\,\bar{u},$$

$$(v_i\Omega_i(k,m-1)+v_j\Omega_j(k,m-1))x\leq v_ib(k,m-1)$$
$$+v_j(k,m-1),$$

$$\forall(i,j)\in P(R(k,m-1)),$$

$$Mx\leq a.$$

The only extra work required is to determine at each step (k,m) the couples (i,j) of rows of this matrix which are in vertex $(R(k,m))$

3.3. *A particular case.* An interesting particular case is when $A^r=\beta I$ for some integer r and some scalar β. Then it suffices to compute the set $R(r,m)$, and the Nr-step controllable set C_{Nr} is the convex polytope defined by

$$|\Omega(r,m)\beta^Nx|\leq\frac{1-\beta^N}{1-\beta}\,b(r,m),$$

and when $|\beta|>1$, the closure of the set of all the controllable states, i.e. the set $C=\bigcup_{N\geq 1} C_N$, is the convex polytope

$$|\Omega(r,m)x|\leq\frac{1}{|1-\beta|}\,b(r,m)$$

When $|\beta| < 1$ (A stable matrix) and the system is controllable, C is the whole space R^n.

4. *Asymptotically controllable set*

In this section we refine some results of the previous section. In most of the practical applications, one wants to design a stabilizing control so that we only need to know the set of states that can be driven to the origin *asymptotically* and not necessarily in *finitely* many steps. With this in mind, we show that we only need to focus on the *unstable part* of the matrix A. We first informally discuss the underlying idea. Without loss of generality, suppose that A is already in the block Jordan decomposition form:

$$A = \left[\begin{array}{c:c} A_1 & 0 \\ \hdashline 0 & A_2 \end{array}\right],$$

where A_1 is a (s, s) matrix with (unstable) eigenvalues of modulus larger than 1 and A_2 is a $(n-s, n-s)$ matrix with (stable) eigenvalues with modulus less than 1. A vector $x \in R^n$ will be written $[x_1, x_2]$ where $x_1 \in R^s$ and $x_2 \in R^{n-s}$. According to (13), the N-step controllable set is the set C_N of vectors $x \in R^n$ which satisfy

$$|\Omega_1(N, m)A_1^N x_1 + \Omega_2(N, m)A_2^N x_2| \leq b(N, m),$$

where $\Omega_1(N, m)$ and $\Omega_2(N, m)$ are submatrices (of appropriate dimension) of $\Omega(N, m)$.

As $N \to \infty$, the set of asymptotically controllable states is unbounded in the subspace spanned by the last $n - s$ columns of A.

To see this, consider a vector $x \in R^n$ which satisfies

$$|\Omega_1(N, m)A_1^N x_1| \leq b(N, m) \tag{14}$$

By definition there is a sequence of controls $u_0, \ldots, u_{N-1}$ such that the initial state $[x_1, 0]$ is driven to zero in N steps. Therefore, given the initial state $x = [x_1, x_2]$ and this sequence of controls, the state z_N after N steps is

$$z_N = A^N[x_1, 0] + \sum_{k=0}^{N-1} A^k B u_k + A^N[0, x_2] = 0 + A_2^N[0, x_2],$$

and thus,

$$\|z_N\| \leq \|A_2^N\| \cdot \|x_2\|.$$

Therefore, any such state can be driven to the origin *asymptotically* if one applies the control $u = 0$ at periods $k \geq N$. Thus, we observe that if one focuses attention on *asymptotic* controllability only, it makes sense to restrict the analysis to the A-unstable subspace. The set of asymptotically controllable states should be a cylinder in R^n and the directions of the cylinder should span the A-stable subspace.

However, to compute the matrix Ω_1, we still need consider the whole system in R^n. The question is, "Can we compute the asymptotically controllable set by just reasoning in the $(n-s)$-dimensional A-unstable subspace?" The answer is yes and therefore great computational savings can be expected. Let us consider the linear system in R^s

$$y_k = A_1 y_{k-1} + LBu_k \qquad k = 0, \ldots, N-1 \tag{15}$$

$$-\bar{u}_k \leq u_k \leq \bar{u}_k, \quad k = 0, \ldots, N-1, \tag{16}$$

where L is the (s, n) matrix $[I_s \mid 0]$ and I_s the (s, s) identity matrix.

This system is the original system where the state $y \in R^s$ is just the first s components of the state $x \in R^n$. Suppose that we compute the N-step controllable set for (15)–(16) defined as

$$C_N^s = \{y \in R^s \mid |\Omega_s(N, m)A_1^N y| \leq b_s(N, m)\}. \tag{17}$$

We have the following result:

Theorem 4.1. For all $N \geq 1$, any initial state $x = [x_1, x_2] \in C_N^s \times R^{n-s}$ can be driven to the origin asymptotically.

Proof. By definition of C_N^s, $x_1 \in C_N^s \Rightarrow$ there exists a sequence of controls $u_0, \ldots, u_{N-1}$ such that after N steps one reaches a state of the form $[0, z]$, $z \in R^{n-s}$. Now at periods $k \geq N+1$, apply the control $u_k = 0$, $\forall k$. Since A_2 has all its eigenvalues with modulus less than 1, starting from a state $[0, z]$ and applying the sequence of controls $0, \ldots, 0, \ldots$, the system is driven asymptotically to the origin. Therefore, all the states $x \in C_N^s \times R^{n-s}$ can be driven asymptotically to the origin. □

Remark 4.1. This set C_N^s is easier to compute than the one described by (14) since it only involves vectors in R^s instead of R^n. It is also larger, because the vectors x_1 satisfying (14) have also the property that the state $[x_1, 0]$ can be driven to the origin in N steps, which is not true in general for a vector $x_1 \in C_N^s$. Of course, since $C_N^s \subset C_{N'}^s$ for all $N' \geq N$, the larger N is the larger is C_N^s.

Ideally one would like to compute $C^s = \overline{\bigcup_{N \geq 1} C_N^s}$. This set C^s is convex but not a polytope in general. Each C_N^s is an inner approximation of C^s. As N increases, more and more bounding hyperplanes are created. C^s is a convex polytope if and only if $A_1^p = \lambda I_s$ for some p and some scalar λ. However, if the smallest eigenvalue (in modulus) is not too close to 1, one can expect that C_N^s is a good inner approximation of C^s for reasonable values of N.

Remark 4.2. If A_1 is a (2, 2) matrix, i.e. if $s = 2$, then a closed-form expression exists for C_N^s and a polynomial time procedure as well (see Lasserre (1991a)).

5. *Stabilizing state-feedback controllers*

In the past few years, some researchers have focused on methods to compute stabilizing state-feedback controllers in the presence of constraints on the controls. In Gutman and Cwikel (1987) for example, some sophisticated nonlinear controllers have been proposed and are computationally involved. However, all the controllable states can be driven to the origin. The linear state-feedback controllers such as those described in Blanchini (1990), Benzaouia and Burgat (1988), Hennet and Beziat (1991), Vassilaki and Bitsoris (1989) and Pajunen and Erdol (1990), are *easier* to build but with a *smaller* controllable region. Some polyhedral set M is chosen and one computes a linear state-feedback controller $u = Fx$ such that M is positively invariant for the closed-loop system $x_{k+1} = (A + BF)x_k$ and is included in the polyhedron $\{x \mid |Fx| \leq \bar{u}\}$. Therefore, the sets C_N built in the previous section could be used to evaluate the quality of the controller F. The set M should be as close as possible (in the Hausdorff metric for example if they are bounded) to C_N for reasonable values of N. Ideally, one would like to compute a state-feedback controller such that C_N is positively invariant.

If the set M is too large, a linear state-feedback controller may not exist. On the other hand, one would like to control states not necessarily close to the origin.

As a possible tradeoff, we propose a simple feedback control law. The idea is to first drive the system to a prescribed relatively small set around the origin and then build a linear state-feedback controller such that this set is positively invariant.

Assume that we have computed *off-line* the set C_N^s for some N. Suppose also that the initial state $x = [x_1, x_2]$ is such that

$$|\Omega_s(N, m)A_1^N x_1| \leq b_s(N, m),$$

i.e. x can be driven to the origin asymptotically (and its projection x_1 in the A-unstable subspace can be driven to the origin in N steps). Driving an initial state to the origin would be achieved in two steps. First, by using a simple feedback controller we drive the projection x_1 of the initial state $[x_1, x_2]$ to a smaller set around the origin, C_1^s for example (or αC_1^s for some positive scalar $\alpha < 1$). At this time the state $z = [z_1, z_2]$ is such that

$$z_1 \in C_1^s \quad \text{and} \quad \|z_2\| \leq K = \|A_2^N\| \quad \|x_2\| + \delta \,.\, S, \tag{18}$$

where $\delta = \max_i | B_i\bar{u}$ and $S = \sum_{j=0}^{\infty} \|A_2^k\|$.

Then, by using Linear Algebra techniques, as in Hennet and Beziat (1991), we build a linear state-feedback controller $u = Fx$ such that the set $\Gamma = C_1^s \times [-K, K]^{n-s}$ is positively invariant by the linear mapping $A + BF$. A nonlinear feedback controller driving an initial state $[x_1, x_2]$ from C_N^s to say C_1^s is described in Lasserre (1991b).

5.1. *A stabilizing linear state-feedback controller.* Now we

assume that the initial state $[x_1, x_2]$ has been driven to a state $[z_1, z_2] \in \Gamma = C_1^s \times [-K, K]^{n-s}$. We want to build a linear state-feedback controller $u = Fx$ such that Γ is positively invariant for the linear mapping $A + BF$ and is included in the set $\{x \mid Fx \leq \bar{u}\}$. Such a controller is proved to be stabilizing (Bitsoris (1988); Hennet and Beziat (1991)). Various techniques have been developed to compute such controllers (Hennet and Beziat (1991); Vassilaki *et al.* (1988); Pajunen and Erdol (1990)). Let us rewrite (18) as

$$\Gamma = \{y \in R^n \mid |Cy| \leq \omega\},$$

where

$$C = \begin{bmatrix} \Omega_s(1, m) & \vdots & 0 \\ \text{—} & \vdots & \text{—} \\ 0 & \vdots & I_{n-s} \end{bmatrix}, \quad \omega = \begin{bmatrix} b_s(1, m) \\ \\ K \end{bmatrix}.$$

Then by Farkas' Lemma (Shriver, 1986), Γ is positively invariant by $A + BF$ if and only if there exists some matrix H_1 such that

$$C(A + BF) = H_1 C \quad \text{and} \quad |H_1|\,\omega \leq \omega,$$

i.e. $|H| - I$ is an M-matrix (Gantmatcher (1959)).

Moreover, to be consistent, the control $u = Fx$ should satisfy $|u| \leq \bar{u}$, i.e. $x \in \Gamma \Rightarrow Fx \leq \bar{u}$. Therefore it suffices to find two matrices H_1, H_2 such that

$$C(A + BH_2C) = H_1 C$$
$$|H_1|\,\omega \leq (1 - \epsilon)\omega$$
$$|H_2|\,\omega \leq \bar{u},$$

for some $\epsilon < 1$, and the linear state-feedback controller is $u = H_2Cx$. This feedback controller can be computed by simple Linear Programming techniques

5.2 *Example*. Let us consider the following system already considered in Hennet and Castelan (1991), but now under the constraints $\pm u_k \leq 1$ on all the controls.

$$x_{k+1} = \begin{bmatrix} -0.219 & 0.383 & 0.529 & 0 & 0.526 \\ 0.047 & 0.519 & -0.671 & 0.687 & -1.003 \\ 0 & 0.831 & 0 & 0.589 & -0.654 \\ -0.679 & 0 & 0.383 & 0.930 & 0.416 \\ -0.934 & 0.053 & 0.067 & -0.139 & 0 \end{bmatrix} x_k + \begin{bmatrix} 0.910 & -0.328 & 0.247 \\ 0.762 & 0.633 & 0 \\ 0.262 & 0.756 & -0.723 \\ 0.047 & 0 & 0.753 \\ -0.736 & 0.365 & 0.651 \end{bmatrix} u_k, \quad k = 0, \ldots$$

After a change of basis to get the block Jordan decomposition of A we finally consider the new system

$$x_{k+1} = \begin{bmatrix} -0.3025 & 1.0279 & 0 & 0 & 0 \\ -1.0279 & -0.3025 & 0 & 0 & 0 \\ 0 & 0 & 0.3455 & 0 & 0 \\ 0 & 0 & 0 & 0.7452 & 0.1402 \\ 0 & 0 & 0 & -0.1402 & 0.7452 \end{bmatrix} x_k + \begin{bmatrix} 0.5130 & -0.0789 & 0.5639 \\ -0.5286 & 0.3307 & 0.0184 \\ -0.2760 & -0.3264 & 1.2683 \\ 0.1382 & 0.0095 & 0.4189 \\ 0.4160 & 0.7176 & -0.7071 \end{bmatrix} u_k, \quad k = 0, \ldots$$

Since there are only two unstable complex eigenvalues, we restrict the analysis to the two-dimensional system

$$y_{k+1} = \begin{bmatrix} -0.3025 & 1.0279 \\ -1.0279 & -0.3025 \end{bmatrix} y_k + \begin{bmatrix} 0.5130 & -0.0789 & 0.5639 \\ -0.5286 & 0.3307 & 0.0184 \end{bmatrix} u_k, \quad k = 0, \ldots$$

We have a two-dimensional system with a (2, 3) control matrix B. In (Lasserre (1991a), a closed-form expression for reachable and controllable sets was derived for two-dimensional systems but with (2, 1) or (2, 2) control matrices. It can be adapted to this case by separating the (2, 3) matrix into a (2, 2) and (2, 1) matrix. Two controllable sets are computed separately for each control matrix and finally the controllable set is the Minkowski sum of the two sets. After some calculations, we get

$$C_1^s = \left\{ x \in R^2 \;\middle|\; \pm \begin{bmatrix} -1.4158 & 2.4703 \\ -5.3713 & 3.0339 \\ 4.4869 & 1.4812 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 2.469 \\ 3.403 \\ 3.8725 \end{bmatrix} \right\},$$

$$C_4^s = \left\{ x \in R^2 \;\middle|\; \pm \begin{bmatrix} 2.5848 & 0.6167 \\ 4.1315 & 4.0096 \\ -0.1289 & -2.4765 \\ 2.5006 & -4.7577 \\ -2.1828 & 0.7699 \\ -4.9173 & -0.9809 \\ 1.2641 & 1.7503 \\ 0.4173 & 4.6606 \\ 1.7734 & 0 \\ -0.0326 & 1 \\ -0.4671 & -1.5873 \\ 0.9037 & -0.2351 \\ -1.2977 & 0.8376 \\ -0.4485 & -0.7467 \\ 1.0916 & 0.9402 \\ -0.5503 & 0.5988 \end{bmatrix} \begin{bmatrix} -0.3025 & 1.0279 \\ -1.0279 & -0.3025 \end{bmatrix}^4 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 12.67 \\ 25.60 \\ 10.23 \\ 24.62 \\ 11.52 \\ 23.92 \\ 9.29 \\ 19.18 \\ 8.82 \\ 4.2 \\ 7.08 \\ 4.63 \\ 7.78 \\ 3.70 \\ 6.55 \\ 3.90 \end{bmatrix} \right\}$$

and a stabilizing linear state-feedback of the form

$$u = Fx = \begin{bmatrix} 0.2137 & -0.8722 & 0 & 0 & 0 \\ 0.7072 & 0.2335 & 0 & 0 & -0.0390 \\ 0.2979 & -0.5198 & 0 & 0 & -0.0481 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}.$$

Starting from an initial state $y_0 = [0.6925, 3.5521, 10, 10, 10]$, where $x_0 = [0.6925, 3.5521]$ is in C_4^s and applying successively the controls $[-1, 1, -1]$, $[-1, 1, 1]$ and $[1; -1; 1]$ one reaches the state $[0.4483, 1.2564, 1.9945, 6.6314, 0.6939]$ where

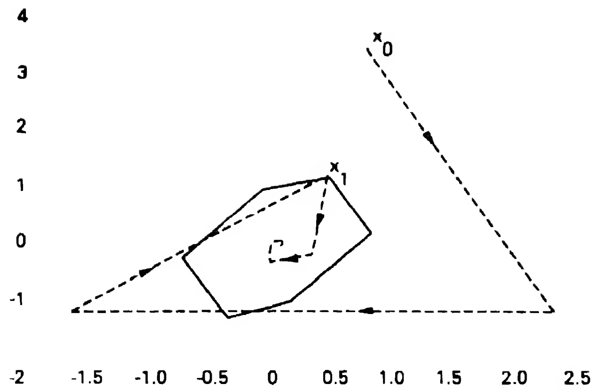


FIG. 1 Example of state trajectory.

$x_1 = [0.4483, 1.2564]$ is a vertex of C_1^u. Then, one applies the linear state-feedback control $u = Fx$. In Fig. 1 the corresponding trajectory is displayed in the A-unstable subspace only. The solid line is the contour of C_1^u and the dashed line is the trajectory.

6. *Conclusions*

In this paper we have presented a simple new method for computing the reachable and controllable sets for input-constrained linear systems. In particular, to characterize the set of initial states that can be driven to the origin asymptotically, it suffices to restrict attention the unstable subspace of the dynamical matrix, thus yielding significant computational savings. A stabilizing feedback controller has also been presented.

References

Barmish, B. R. and W. E. Schmitendorf (1980). A necessary and sufficient condition for local constrained controllability of a linear system. *IEEE Trans. Aut. Contr.*, **25**, 97–100.

Benzaouia, B. and C. Burgat (1988). Regulator problem for linear discrete time systems with nonsymmetrical constrained control. *Int. J. Contr.*, **48**, 2441–2451.

Bitsoris, G. (1988). Positively invariant polyhedral sets of discrete-time linear systems. *Int. J. Contr.*, **47**, 1713–1726.

Blanchini, F. (1990a). Feedback control of linear time invariant systems with state and control bounds in the presence of disturbances. *IEEE Trans. Aut. Contr.*, **35**, 1231–1234.

Blanchini, F. (1990b). Constrained control for perturbed linear systems. *Proceedings of the 29th Conference on Decision and Control*, Honolulu, HI.

Bolker, E. (1969). A class of convex bodies. *Trans. Am. Math. Soc.*, **145**, 323–345.

Desoer, C. A. and J. Wing (1961). The minimal time regulator problem for linear sampled-data systems: general theory. *J. Frank. Inst.*, **272**, 208–228.

Fisher, M. E. and J. E. Gayek (1990). Estimating reachable sets for two-dimensional linear discrete systems. *J. Opt. Th. Appl.*, **56**, 67–88.

Gantmatcher, F. R. (1959). *The Theory of Matrices*. Chelsea, New-York.

Gutman, P. O. and M. Cwikel (1987). An algorithm to find maximal state constrained sets for discrete-time linear dynamical systems with bounded controls and states. *IEEE Trans. Aut. Contr.*, **32**, 251–254.

Hamza, M. and M. E. Rasmy (1971). A simple method for determining the reachable set for linear discrete systems. *IEEE Trans. Aut. Contr.*, **16**, 281–282.

Hennet, J. C. and J. P. Beziat (1991). A class of invariant regulators for the discrete-time linear constrained regulation problem. *Automatica*, **27**, 549–554.

Hennet, J. C. and E. B. Castelan (1991). A unified framework for linear constrained regulation problems. *Proceedings IMACS Symposium MCTS*, Casablanca, Morocco.

Lasserre, J. B. (1987). A complete characterization of reachable sets for constrained linear time-varying systems. *IEEE Trans. Aut. Contr.*, **32**, 836–838.

Lasserre, J. B. (1991a). On reachable and controllable sets for two-dimensional linear discrete-time systems. *J. Opt. Th. Appl.*, **70**, 583–595.

Lasserre, J. B. (1991b). Reachable, controllable sets and stabilizing control of constrained linear systems. LAAS-report 91-311.

Lin, J. N. (1970). Determination of the reachable set for a linear discrete system. *IEEE Trans. Aut Contr*, **15**, 339–342.

Pajunen, G. and N. Erdol (1990). Time varying saturating feedback control of linear systems under state and control constraints. *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI.

Rumchev, V. G. and D. J. G. James (1989). Controllability of positive linear discrete-time systems *Int. J. Contr*, **50**, 845–857.

Schrijver, A. (1986). *Theory of Linear and Integer Programming*. John Wiley, New York.

Van Til, R. and W. E. Schmitendorf (1986). Constrained controllability of discrete-time systems. *Int. J. Contr.*, **43**, 941–956.

Vassilaki, M. and G. Bitsoris (1989). Constrained regulation of linear continuous time dynamical Systems. *Syst. & Contr. Lett.*, **13**, 247–252.

Vassilaki, M., J. C. Hennet and G Bitsoris (1988) Feedback control of linear discrete time systems under state and control constraints. *Int. J. Contr.*, **47**, 1727–1735.

Witsenhausen, H. S. (1972) A remark on reachable sets of linear systems. *IEEE Trans. Aut. Contr.*, **27**, 420–422.

Automatica, Vol. 29, No. 2, pp. 537–539, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

Necessary and Sufficient Conditions for Global Optimality for Linear Discrete Time Systems*

N. S. ROUSAN†

Key Words—Optimal control; system theory; discrete time systems; optimal regulators; suboptimal control.

Abstract—This paper extends the results introduced by Huang and Li (1989, *Int. J. Control*, **50**, 2341–2347) to discrete time systems. Necessary and sufficient globally optimal conditions, in the form of a matrix equation and a matrix inequality, are presented for the existence of the optimal constant output feedback gain of discrete time invariant system. Furthermore, it is shown that if the optimal output gain L_0 exists, it must satisfy $L_0 C_0 = K_0$ where K_0 is the optimal state feedback gain. An example is given to show that a globally optimal output law may not be found even if the system is stabilizable by output feedback (i.e. $\Re(K_0^T) \not\subset \Re(C^T)$). These results shed some light on the fundamental issues of output feedback and justify suboptimal solutions.

1. *Introduction*

CONSIDER THE FOLLOWING discrete time invariant control system.

$$x_{k+1} = Ax_k + Bu_k, \tag{1}$$

$$y_k = Cx_k, \tag{2}$$

where $x_k \in \Re^n$ is the state vector; $u_k \in \Re^m$ is the control vector; and $A \in \Re^{n \times n}$, $B \in \Re^{n \times m}$, and $C \in \Re^{l \times n}$ are constant matrices. Assume the system given above is stabilizable and the matrix C has a full-row rank (i.e. rank $(C) = l$). Let the performance criterion have the quadratic form

$$J = \sum (x_k^T Q x_k + u_k^T R u_k), \tag{3}$$

where $R \in \Re^m$ symmetric positive definite matrix, $Q = F^T F$ and (A, F) is detectable.

Concerning the above optimal control problem, it is known that there are two kinds of optimal constant feedback problems.

Problem 1. Let $J(x_0, Kx_k)$ denote the cost (3) when the control takes the state variable feedback form $u_k = Kx_k$ and the initial state is x_0. Consider finding the optimal constant feedback gain $K_0 \in \mathbb{K}$ so that

$$J(x_0, K_0 x_k) \le J(x_0, Kx_k), \quad \forall x \in \Re^n \quad \forall K \in \mathbb{K},$$

where $\mathbb{K} = \{K : \sigma(A + BK) \subset \mathbb{D}_1\}$ in which $\sigma(\cdot)$ denotes the spectrum of a square matrix and $\mathbb{D}_1$ the unit disk in the complex plane.

Problem 2. Let $J(x_0, Ly_k)$ denotes the cost (3) when the control takes the output feedback form $u_k = Ly_k$ and the initial state is x_0. Consider finding the optimal constant feedback gain $L_0 \in \mathbb{L}$ so that

$$J(x_0, L_0 y_k) \le J(x_0, Ly_k), \quad \forall x \in \Re^n, \quad \forall L \in \mathbb{L},$$

where $\mathbb{L} = \{L . \sigma(A + BLC) \subset \mathbb{D}_1\}$.

It is well known that $u_k^* = K_0 x_k$ represents the optimal control law for Problem 1 where K_0 is the solution of an algebraic Riccati equation. Due to the impressive approaches to the matrix Riccati equation, the existence and uniqueness problems have been solved completely (Dorato and Levis, 1971; Franklin and Powell, 1980).

However, it is often the case that a complete set of state variables is not directly available for feedback purposes. Therefore, the output feedback problems such as stabilization, pole assignment, and optimal control have greater value than complete state variable feedback problems from the practical viewpoints.

The optimal output feedback problem has received great attention. Direct minimization of criterion (3) would produce feedback gains that are dependent upon the initial state x_0. In order to eliminate this dependency, it was proposed (O'Reilly, 1980) that the expected value of criterion (3) be minimized assuming that the initial state is normally distributed with $E\{x_0\} = 0$. Therefore already obtained results are suboptimal instead of globally optimal. Until now the fundamental problems on optimal output feedback have not been answered and still occupy control system researchers. These fundamental issues are as follows.

(a) The existence and uniqueness of an optimal control solution when the system is stabilizable by output feedback.

(b) The relationship between the optimal law of Problem 2 and that of Problem 1.

This paper studies the above issues which relate to Problem 2. The contribution can be summarized as answering the existence and uniqueness problem, showing the relationship between the optimal output feedback gain L_0 and the optimal state variable feedback gain K_0; that is, if L_0 exists, there must be $L_0 C = K_0$. The technique in the present paper more closely follows that of the continuous-time system counterpart of Huang and Li (1989).

This paper is organized as follows. In Section 2 the global optimality conditions to Problem 2 are applied and the necessary and sufficient conditions of existence are derived. The main Theorem 2 is introduced in Section 3. An example is presented in Section 4. It shows that the system, in some cases, may not have an optimal output feedback law even if it is stabilizable by output feedback.

2. *A necessary and sufficient condition of existence*

Proposition. Consider the system (1), (2) and (3). If the control $u_k = LCx_k$ such that $A + BLC$ is stable, then the performance index (3) can be rewritten as follows:

$$J = x_0^T P x_0 + \sum_{k=0}^{\infty} x_k^T[(A + BLC)^T P(A + BLC) + Q + (LC)^T R(LC) - P]x_k, \tag{4}$$

where $P \in \Re^{n \times n}$ is a positive definite matrix.

* Received 30 July 1991; revised 10 March 1992; received in final form 7 May 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor P. Dorato under the direction of Editor H. Kwakernaak.

† Mu'tah University, Electrical Engineering Department, Karak, Jordan.

Proof. Noting that for a stable system $x_\infty = 0$, observe that

$$\sum_{k=0}^{\infty} (x_k^T P x_k - x_{k+1}^T P x_{k+1}) = x_0^T P x_0.$$

We can therefore add zero, in the form of the left-hand side of the above equation minus its right-hand side, to the performance index (3). Also, by substituting $u_k = LCx_k$, we get (4). □

The equation can be written as the perfect square of a norm with respect to $(B^T PB + R)$ (McReynolds, 1966):

$$J = x_0^T P x_0 + \sum_{k=0}^{\infty} \|[(B^T PB + R)^{-1} B^T PA + LC]x_k\|^2_{(B^T PB + R)}. \tag{5}$$

Notice that the minimum cost (5) is reached when the summand is zero (assuming $L = L_0$ is the optimal output gain). In other words, when the matrix equation

$$(A + BL_0C)^T P(A + BL_0C) + Q + (L_0C)^T R(L_0C) - P = 0. \tag{6}$$

Theorem 1. $J = x_0 P_0 x_0$ and $u_k = L_0 y_k$ are the minimum cost and optimal control, respectively for Problem 2 if and only if the following hold:

(a) $(A + BL_0C)^T P_0(A + BL_0C) + Q + (L_0C)^T R(L_0C) - P_0 = 0, \quad (7)$

(b) $(A + BLC)^T P_0(A + BLC) + Q + (LC)^T R(LC) - P_0 \geq 0, \quad \forall L \in \mathbb{L}. \quad (8)$

The proof of Theorem 1 follows from equations (4) and (5).

3. *Main result*

In order to derive our main Theorem 2, we first need to introduce a lemma. This lemma has been proven by Huang and Li (1989).

Lemma 1. For any matrix $S \in \Re^{n\times m}$, $S \neq 0$, and any full row rank matrix $C \in \Re^{l\times n}$, there exists a matrix $X \in \Re^{m\times l}$, $\|X\|_F = 1$, a vector $v \in \Re^n$, $\|v\|_2$, and $\alpha > 0$ such that

$$v^T [SXC + (SXC)^T] v < -\alpha, \tag{9}$$

where $\|\cdot\|_F$ is the Frobenius norm of the matrix.

Definition. The optimal output feedback gain L_0 is said to be a derivative solution of the corresponding optimal state variable feedback problem if $L_0C = K_0$, where K_0 is the optimal feedback gain for the latter problem.

Theorem 2. The optimal output feedback gains are all derivative solutions of the corresponding optimal state variable feedback gain.

Proof. Let P_0 and L_0 be the optimal solutions of Problem 2, so that P_0 and L_0 satisfy (a) and (b) of Theorem 1.

Consider $L_0(\epsilon, X) = L_0 + \epsilon X$, $\|X\|_F = 1$. Then there exists $\epsilon_0 > 0$ such that $L_0(\epsilon, X) \in \mathbb{L}(A, B, C)$, for $|\epsilon| < \epsilon_0$. This is true because $\mathbb{L}(A, B, C)$ is an open set and $L_0 \in \mathbb{L}(A, B, C)$. Therefore

$$[A + B(L_0 + \epsilon X)C]^T P_0[A + B(L_0 + \epsilon X)C] + Q + [(L_0 + \epsilon X)C]^T R[(L_0 + \epsilon X)C] - P_0 \geq 0. \tag{10}$$

Using (7), (10), can be rewritten as

$$\epsilon^2 (XC)^T [R + B^T P_0 B](XC) + \epsilon[[(R + B^T P_0 B)L_0C + B^T P_0 A]^T (XC) + (XC)^T [(R + B^T P_0 B)L_0C + B^T P_0 A]] \geq 0. \tag{11}$$

If $S = [(R + B^T P_0 B)L_0C + B^T P_0 A]^T \neq 0$, from Lemma 1, there exists $X \in \Re^{n\times m}$ with $\|X\|_F = 1$ and $v \in \Re^n$ with $\|v\|_2 = 1$ so that

$$\epsilon v^T [SXC + (SXC)^T] v < -\epsilon\alpha, \quad \alpha > 0.$$

Since $v^T [(XC)^T [R + B^T P_0 B](XC)] v$ is finite, then the following holds:

$$\epsilon^2 v^T [(XC)^T [R + B^T P_0 B](XC)] v - \epsilon\alpha < 0, \tag{12}$$

for ϵ small enough. This contradicts equation (11) which shows that S must be a zero matrix, i.e.

$$S = [(R + B^T P_0 B)L_0C + B^T P_0 A]^T = 0.$$

Furthermore,

$$L_0C = -(R + B^T P_0 B)^{-1} B^T P_0 A = K_0. \tag{13}$$

□

From the above main theorem, we can see that finding a globally optimal output feedback law reduces to the following procedure.

(a) Obtain P_0 and K_0 by solving the discrete Riccati equation; if $\Re(K_0^T) \subset \Re(C^T)$, then there exists L_0 so that $L_0C = K_0$, i.e. Problem 2 is solvable, otherwise unsolvable.

(b) $\Re(K_0^T) \subset \Re(C^T)$ if and only if

$$\mathbb{N}(C) \subset \mathbb{N}(K_0) = \mathbb{N}((R + B^T P_0 B)^{-1} B^T P_0 A) = \mathbb{N}(B^T P_0 A) = A^{-1} P_0^{-1} \mathbb{N}(B^T),$$

or equivalently

$$P_0 A \mathbb{N}(C) \subset \mathbb{N}(B^T). \tag{14}$$

This means that if $P_0 A \mathbb{N}(C) \subset \mathbb{N}(B^T)$, Problem 2 is solvable otherwise unsolvable.

As we know, P_0 determined by the discrete Riccati equation depends on the given A, B, Q, and R. In general cases condition (14) can hardly be satisfied. In the particular case when rank $(B) = m - l =$ rank (C), for instance, (14) is never satisfied. To show that Problem 2 is nearly unsolvable even if the system is stabilizable by output feedback, we present the following example.

4. *An example*

Consider the following discrete linear system:

$$x_{k+1} = [A - blc]x_k,$$

where

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad c = [0 \quad 1].$$

The system is controllable and observable. The characteristic polynomial $P(\lambda)$ can be written as

$$P(\lambda) = |\lambda I - (A - blc)| = \lambda^2 - \lambda + l,$$

we notice that if $l = \frac{1}{4}$, the roots of $P(\lambda)$ will be within the unit circle. This means the system is stabilizable by output feedback.

Consider the performance criterion

$$J = \sum^{\infty} (x_k^T Q x_k + u_k^T u_k),$$

with

$$Q = \begin{bmatrix} \delta^2 & \mu \\ \mu & \beta^2 \end{bmatrix}, \quad R = 1.$$

In this case, the discrete Riccati equation

$$P = Q + A^T P(I + BR^{-1} B^T P)^{-1} A,$$

is equivalent to the following scalar equations

$$p_{11} = \delta^2 - \frac{p_{12}^2}{1 + p_{11}} + p_{22},$$

$$p_{12} = \mu - \frac{p_{12}^2}{1 + p_{11}} + p_{22},$$

$$p_{22} = \beta^2 - \frac{p_{12}^2}{1 + p_{11}} + p_{22},$$

where

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}.$$

The optimal state variable feedback gain is

$$K_0 = -(R + B^TPB)^{-1}B^TPA = -\frac{1}{1+p_{11}}[p_{12} \quad p_{12}].$$

Obviously, the globally optimal output feedback problem is unsolvable even though the system is stabilizable by output feedback. This example shows that $\mathfrak{R}(K_0^T) \not\subset \mathfrak{R}(C^T)$, or there exists no output gain L such that $LC = K_0$.

5. *Conclusions*

The fundamental theorem on the optimal output feedback problem with quadratic performance criterion is given. The theorem shows that if there exists a globally optimal output feedback law L_0, then it must satisfy $L_0C = K_0$, where K_0 is the unique solution of the discrete Riccati equation. Moreover, an example to show that the condition (14) is hardly satisfied even if the system is stabilizable. These give the reasons for finding suboptimal laws that stabilize the system. As far as output feedback problems such as stabilization, pole assignment, and optimal control are concerned, there are still many questions awaiting investigation.

References

Dorato, P. and A. H. Levis (1971). Optimal linear regulators: the discrete-time case. *IEEE Trans. Aut. Control,* **16,** 613–620.

Franklin, G. F. and J. D. Powell (1980). *Digital Control of Dynamic Systems.* Addison-Wesley, MA.

Huang, L. and Z. Li (1989). Fundamental theorem for optimal output feedback problem with quadratic performance criterion. *Int. J. Control,* **50,** 2341–2347.

McReynolds, S. R. (1966). Ph.D thesis. Harvard University, Cambridge, MA.

O'Reilly, J. (1980). Optimal low-order feedback controllers for linear discrete time systems. *Contr. Dynam. Syst.,* **16,** 325–367.

Automatica, Vol. 29, No. 2, pp. 541–548, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Brief Paper

The Safety of Process Automation*

A. TOOLA†

Key Words—Control system design; process control; reliability; automation.

Abstract—The effect of automation on process safety is not clear. On the one hand, automation is blamed for posing risk and for increasing the chance of human error in situations involving disturbances; on the other hand, it is admitted that automation enables sophisticated process control and handling of disturbance situations without human interference. The methods of safety analysis can be applied during the designing stages of safe process automation. The hazard and operability study makes it possible to take into account the potential process disturbances and to develop countermeasures for them. Action error analysis studies the consequences of potential human errors in task execution. Fault tree analysis can be used to study the causes of potential accidents and to examine the control actions suitable for providing protection against them thereby reducing the probability of accidents. Event tree analysis is a method for considering the consequences of potential hazardous situations and for developing countermeasures to reduce such consequences. Failure mode and effect analysis is a method for checking that the potential failures of the control and automation system are not overlooked. Reliability assessment can be used with safety analysis methods to study the bottlenecks in the design and to prioritize the countermeasures whereby the risk can be reduced to attain an acceptable level

Introduction

THERE ARE methodologies and procedures available for analysing hazards in process industry, but the safety of computer-controlled systems is not as well established. This paper examines whether the methods of conventional safety analysis can be applied to the safety design of a computer controlled system. To start from the beginning, the definitions for "safety" and related concepts given by various sources are studied, and shown in Table 1.

Safety describes the non-existence of risk posed by a system, or to the system. The basic elements of safety are the causes and the consequences of accidents. In the definitions of safety, the consequences are defined to be the effects on human life, property or the environment.

According to the Shen's definition (Shen, 1986), safety is a measurable concept. In any case, it is possible to say whether the system is in a safe state or an unsafe state. Another question is whether, or not, it is possible to define if a system will be safe enough throughout its whole life-cycle.

In some of the definitions safe means zero-risk (under defined conditions): safe enough is a condition where there is no possibility of an accident. In many cases this would mean that the intended function of the process plant is impossible. For example, the risk posed by a chemical plant will in most of the cases be at its minimum when the plant is not operating, but this would not be reasonable for the production target of the plant. A more realistic safety policy is risk minimization or a risk optimization target, under certain constraints. This includes the aspect of concentrating on the weak points of the process in order to make them safer. An even more sophisticated approach would be to minimize the total cost of the plant during its life-cycle. This total cost consists of, e.g. the costs of design, start-up, operation and maintenance (both preventive and corrective maintenance), unavailability costs and costs of accidents.

When the causes of accidents are discussed, the concepts of failure, fault, error and mistake are often mentioned. System failure is defined as the inability of a system to fulfil its operational requirements. Systematic failures cause the system to fail under some particular combinations of inputs or under some particular combinations of inputs or under some particular environmental condition. Systematic failures could arise at any part of the safety life-cycle. Other types of failure are those whose occurrence follows a stochastic model. Examples are failures due to ageing of mechanical components, and random failures of electronic components. Fault and error are defined as reasons for failure. A mistake is defined as a human error or fault

The concept of hazard comes up frequently in accident research. It describes the potential of accidents. The concept of an accident or a mishap is rather vague. It lies between the causes and consequences, but the limit where the accident event begins and where it ends, i.e. which are the causes and which the consequences, is not exact.

In conclusion, the above definitions make it possible to derive the following: causes of accidents can be divided into systematic failures and stochastic failures in systems. A software failure is always a systematic failure. Examples of other systematic failures are, e.g. those caused by errors and mistakes in specification, design, construction, operation or maintenance. The occurrence of failure does not cause the accident directly. Typically, a specific state of the process, or a combination of failures, are needed for a hazard to occur and to develop to be an accident (Fig. 1). A viewpoint is that all the accidents are caused by a human being, either by the design organization which has not designed the system to be safe enough, or by the operational organization which has not been able to handle the disturbance situations. Human aspects in accidents are studied, for example, in Rasmussen (1986, 1990), Rasmussen *et al.* (1987) and Reason (1990a,b).

An automation system itself is rather safe. No great amount of energy is bound to it. The hazards arise in the process controlled by the automation. The safety of a process plant is based on the process design, which defines the number of possible unsafe states and their probability. However, responsibility for many safety features is allocated to the automation design. In a hazardous situation, the target of the control system is to prevent an accident from happening by keeping the process in a safe state or by bringing the process back to a safe state before any serious consequences have occurred. Sometimes it may not be possible to maintain a safe state; in such case, the control system should minimize the consequences. Another important safety-related target of the automation design is to prevent hazardous situations from being caused by automation (Fig. 2).

* Received 8 February 1991; revised 11 July 1991; revised 9 January 1992; revised 11 April 1992; received in final form 9 June 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Editor A. P. Sage

† Technical Research Centre of Finland, Safety Engineering Laboratory, P.O. Box 656, 33101 Tampere, Finland.

TABLE 1. DEFINITIONS FOR SAFETY AND RELATED CONCEPTS

Term	Definition
Safety	Freedom from those conditions that can cause death, injury, occupational illness, or damage to or loss of equipment or property (MIL, 1984; IEC, 1989a,b). Safety is a measure of the degree of freedom from risk in any environment (Leveson, 1986). Safety of a system is the probability that, when operating and/or residing under stated conditions, the system will not be injured significantly for a specified interval of time (Shen, 1986).
System failure	A system failure occurs when the delivered service deviates from the specified service, where the service specification is an agreed description of the expected service. A failure, in short, is the manifestation of an error on the system or software (IEC, 1989a,b).
Systematic failure	Failures due to errors (including mistakes and acts of omissions) in design, construction or use of a system which cause it to fail under some particular combinations of inputs or under some particular environmental condition (IEC, 1989b).
Fault, error	The cause of an error is a fault which resides, temporarily or permanently, in the system. An error is that part of the system state which is liable to lead to failure (IEC, 1989a,b).
Mistake	A human error or fault. A human action (in carrying out any system life-cycle activity) that may result in failure (IEC, 1989a,b)
Hazard	A condition that is prerequisite to a mishap (MIL, 1984). A physical situation with a potential for human injury, damage to property, damage to the environment of some combination of these (Anon., 1985).
Accident, mishap	An unplanted event or series of events that results in death, injury, occupational illness, or damage to or loss of equipment or property (MIL, 1984).
Unsafe-state	A state that have been determined to be unacceptable in that their risk exceeds a specified threshold (Leveson, 1984).
Safe-state	A state when the specified hazard no longer exists, or, a state of a defined system in which there is no danger to human life, economics or environment under certain assumptions and specified conditions (IEC, 1989a).

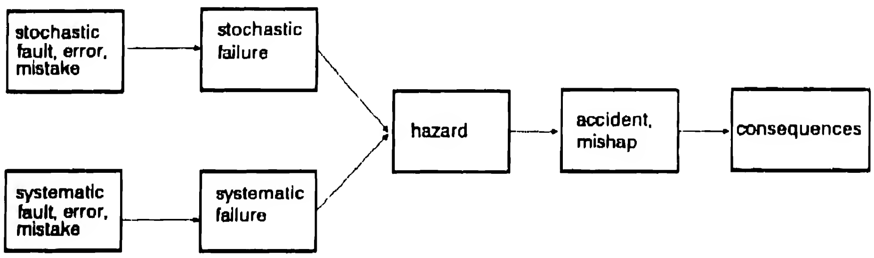


FIG. 1. The accident is caused by unintended events in human activity, in the technical equipment or in the environment

In practice, safety design consists of identifying possible process disturbances and potential accidents, whereafter preventive measures are designed for them. Some rules of thumb are used to define what is safe enough and what is not. In some cases, and for restricted problems, more sophisticated methods and safety measures are used for decision-making.

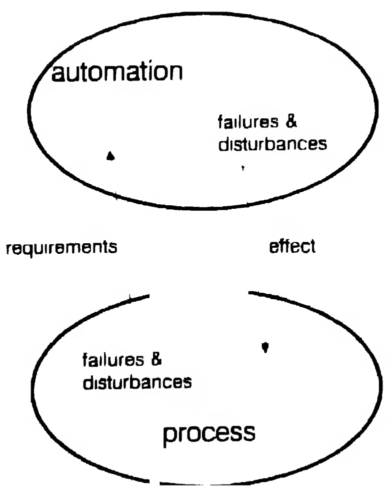


FIG. 2. The safety of the process automation is closely linked with the safety features of the process itself.

When designing the process automation, the designer has the following options for reducing the risk related to a specific accident (see Fig. 3).

Firstly, potential causes of an accident which exist in automation can be eliminated or their probability minimized. This measure includes reduction of the number of potential systematic failures. It can be accomplished, for instance, by designing procedures, through management actions, by having skilled and experienced designers and by thorough testing. Stochastic failures can be reduced by using components of adequate quality, by ageing procedures and by testing procedures for stand-by components. Redundancy, both functional and architectural, can be employed on different levels, e.g. on component, circuit-board, station and system levels. Secondly, it is possible to prevent the unsafe state from occurring by designing countermeasures for hazards by which the system can be brought back into a safe state. This can be done, for example, by designing alarms and proper operator actions for them, by designing automatic control actions, by designing fail-safe systems, interlocks and trip-systems, and through building redundancy and stand-by systems. Thirdly, it may be possible to control the unsafe state or the consequences of the hazard, and to minimize them, i.e. to keep the process in a state which, though perhaps unsafe, still prevents any more harm from occurring. This can be done, e.g. by means of safety protection equipment or by keeping people and materials out of the hazardous area.

Accidents caused by automation systems

This chapter represents some accident studies of process industries, and interprets them from the automation point of

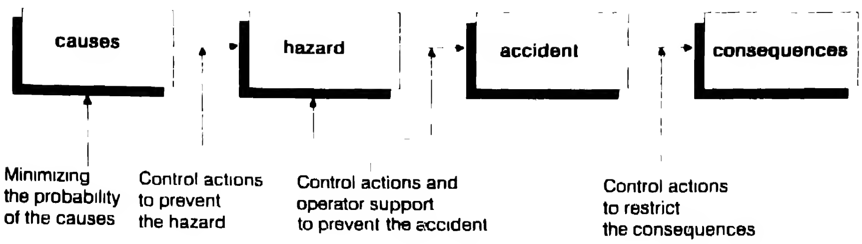


FIG. 3. The automation engineer can affect the probability and severity of accidents in many ways.

view. The purpose is to get an overview of the problems and of the safety studies which are needed to detect the potential causes of mishaps before their occurrence.

The VTT's safety engineering laboratory has studied 31 accidents or disturbances in the process industry. Most of the accidents were selected from the VARO-register, which is comprised of accidents which occurred in Finland. In addition, some accidents reported in the literature (Kletz, 1983; Lees, 1980) and one from the FACTS database were studied. The first criterion for inclusion in the study was that the automation or control system had contributed to the accident. Another criterion was that either the process was a batch process or the accident took place in a sequential operation in a non-batch process. Such operations are, for example, the start-up or shut-down of continuous processes.

The purpose of the study on accidents was to identify ways to improve the safety design of sequence logic functions. Owing to the assumption that discontinuous functions are both more complex to control than continuous ones (Rosenof and Ghosh, 1987) and more critical to safety (Rasmussen, 1986), the results of the study on accidents can be used to get an idea of the accident causes and, thus, clarify requirements for safety analysis and design of process automation.

The accidents were classified according to the status of the process at the moment when the disturbance took place, the cause of the failure, and according to the role of an interlock system in the accident (Table 2) In most cases there were at least two causes. For example, if in the design the possibility of an equipment failure had not been considered, both a design error and an equipment failure were thought to be the causes of the accident. Thirteen of the cases had no interlock system, but after the accident it was concluded that such a system would have been easy to realize, feasible and would have prevented the accident. In three of the cases there was an interlock system which, however, had been overridden. Furthermore, in three of the cases the interlock had failed to function.

TABLE 2. THE RESULTS OF THE STUDY ON 31 ACCIDENTS BY VTT AND THE STUDY ON 17 INCIDENTS BY PITBLADO *ET AL.* (1989). THE INCIDENTS WERE CAUSED BY FAILURES IN AUTOMATION AN INCIDENT IS HAVING MANY CAUSES WHICH MAKES THE TOTAL NUMBER OF CAUSES EXCEED THE NUMBER OF SAMPLES

State of the unit:	
Commissioning	1
Start-up	7
Shut-down	0
Maintenance	5
Normal operation	18
Cause of the accident:	
Design error	23
Installation error	1
Operator error	11
Equipment failure	12
Role of the interlock during the accident	
Did not exist	13
Was overridden	3
Failed to function	3

One result of the study was the conclusion that possible equipment failures and their consequences should be considered systematically when the control or automation system is being designed. In this way the control system will be able to handle the situation in case of failures in process or control system equipment. Correspondingly, the possibility of operator errors should be taken into account. Especially human errors in disturbance and failure situations should be studied. The trips and interlocks should be developed so that they do not disturb the operation of the unit in any situation which could be considered to be normal and safe. When a solution has been found that is acceptable with regard to both availability and safety, operators need not disable the protective function.

Pitblado *et al.* (1989) studied 17 hazards which occurred over a four year period in computer controlled plants in The Netherlands. The hazards were related to the computer system or to human interaction. Hardware failure was involved in nine of the hazards Software failure was recognized in three of the hazards. one failure in system software and two in site software implementation. Human errors were classified according to the life-cycle of the process. Five design errors were found, two installation errors, one in the commissioning and testing phase, 10 in operations and two in maintenance. The most common type of error (accounting for 10 errors in all) was that due to inadequate, insufficient or incorrect information supplied to the person(s) involved. Recognition failure, despite an adequate supply of information, took place in the case of two hazards. Failure to follow procedures correctly was another type of error. Eight such errors were involved in the hazards.

Several other studies have investigated accident causes in process industry in general, e.g. the works of Rasmussen (1986) and Pipatti (1989). In the following a brief overview is given about what they found with respect to the automation or control system (Table 3). Rasmussen studied the accidents which involved hazardous chemical reactions. The study was comprised of 190 accidents, and the factors which had caused the accidents were classified on many levels. The causes related to automation or control were as follows: poor instrumentation was identified in 19% of the incidents; operator error was likewise involved in 19% of the hazards. Poor instrumentation included inadequate alarms, interlock systems, control systems, etc. Such incidents may have resulted from a variety of reasons: some failure of equipment, incorrect or unsatisfactory design of the control or protective system, inadequate inspection, or some human error by the operator. According to Rasmussen, poor instrumentation, insufficient design and operator error were noted very often in combination with batch operations.

In the accident study of Pipatti (1989), failures in instrumentation were present in about 5–15% of all accidents (depending on the data sources). Human errors (about 17–26% of all the accidents) had been caused mainly by negligence or ignorance. For example, notification of changes which had been made in the process was not given to everyone who needed that information. Typically, the human errors were made during the start-up of the process. In some cases, the consequences of the human errors might have been smaller had the degree of automation been higher.

The accident reports and statistics reveal that the automation or control system as a cause of accidents has not

TABLE 3. SUMMARY OF THE ACCIDENT STUDIES OF RASMUSSEN AND PIPATTI

Reference	Sample	Automation related causes of the accidents. Poor instrumentation.	Human error.
Rasmussen (1986)	190 accidents which involved hazardous chemical reactions.	19%	19% (only operator errors)
Pipatti (1989)	84 accidents which happened in Finland.	5%	26%
	184 accidents which happened outside Finland.	15%	17%

been studied in detail. It is hard to find statistics on the effect of automation failures or automation deficiencies on the availability or safety of process plants. One reason for this lack may be that accident reports do not consider automation as a "unit" itself, but rather a part of the process unit controlled by the automation.

On the basis of the accident studies mentioned, safety design of process automation should take the following aspects into account. Firstly, on the basis of Rasmussen and Pipatti, the co-operation between automation and human operator is important to avoid human errors during operation, especially in case of discontinuous process functions. Secondly, the occurrence of equipment failures as causes of accidents bring up the requirement that potential failures, both in measurement and control equipment, and in process equipment, should be studied, and the design should prepare for them, an equipment failure of the system should not lead to an accident. In the same way, potential human errors should be recognized and prepared for. A key feature here is the quality of the process information provided to the operator.

Design procedures for a safe automation system

Roughly, the safety design should involve the following steps.

Firstly, the elicitation of requirements is performed. The safety targets of the automation should be derived from the safety features of the process, and the requirement elicitation

TABLE 4. SAFETY ANALYSIS METHODS SCOPE AND PRINCIPLE

Analysis method	Purpose, scope	Principle	References
Hazard and operability study (HazOP)	HazOp is widely used for hazard identification in process industry in order to discover potential hazards and operability problems.	HazOp studies the potential deviations from the intended operation conditions. The studies are carried out by a multidisciplinary team, and key words are used to guide the analysts.	Anon. (1977) Lees (1980) Kletz (1983a) Nimmo *et al.* (1987) Pitblado (1989) Wells (1980)
Action error analysis (AEA)	AEA considers the operational, maintenance, control and supervision actions performed by a human being. The potential mistakes in individual actions are studied.	A short checklist is used. The effects of each potential mistake on safety and on system performance, recognition of the mistake, and potential countermeasures are planned	Taylor (1981)
Fault tree analysis (FTA)	FTA models the cause sequence leading to the TOP-event. FTA can be used as a quantitative method.	The causes are modelled backwards, and the probability of the TOP-event is assessed on the basis of this model and reliability figures of the system components.	IEC 1025 (1990) Wells (1980) Lees (1980) Anon. (1989) Malasky (1982)
Event tree analysis (ETA)	ETA models the potential consequence sequences of a hazardous situation or event. ETA can be used as a quantitative method.	ETA reasons forwards starting from the hazardous situation, and the potential consequences are modelled.	Wells (1980) Anon. (1989) Suokas (1988)
Failure mode and effect analysis (FMEA)	In FMEA the possible failures of the system components or subsystems and the consequences are analysed systematically. FMEA is commonly used for mechanical, electrical and electronic components.	The components of the system and their failures and failure modes are listed on a tabular sheet. Checklists can be used to support the analysis.	Wells (1980) Malasky (1982) IEC 812 (1985) Lehtelä (1990)
Reliability assessment	Reliability assessment means quantitative studies on potential component and equipment failures, their causes and consequences.	Block diagram or FTA is used as a basis. The most important measures in reliability assessment are failure rate and time concepts.	MIL 217 (1986) Lees (1980) Anon. (1989) Dhillon and Singh (1981) O'Connor (1981) Toola (1988)

should be based on studies of the safety aspects of the process. The questions of "safe enough" and acceptable risk must be answered. The safety requirements may be either qualitative or quantitative. Safety may be measured in terms of the probability and severity of accidents, or it may be agreed to be achieved by using certain standards and design practices. In specifying the requirements, it should be remembered to specify them in a way that makes verification possible.

Secondly, the design of the automation takes place. During this planning the safety aspects are designed. It is much easier to design them in the early phase than to make extensive modifications to the design later (Watson, 1989).

Finally, the safety of the automation is verified. The verification includes both the functional and non-functional features. The verification methods depend on how the specification is expressed.

The design of safety critical control systems has been illustrated and guidance given, e.g. in Anon. (1986, 1987a, b, c) and Bloomfield and Brazendale (1990).

This chapter studies some safety analysis methods in automation design. The analyses are introduced in Table 4 where their scope and principle are shown. References for further information are also given. Table 5 introduces the methods from an automation design point of view: the advantages of the methods and their usefulness in control

TABLE 5. SAFETY ANALYSIS METHODS. ADVANTAGES, RESTRICTIONS AND DEFICIENCIES

Analysis method	Advantages in automation design	Restrictions and deficiencies in automation design
Hazard and operability study (HazOp)	The outcome is a list of action, e.g. design changes for consideration, cases identified for more detailed study, etc. The contribution for automation design are the countermeasures related to the instrumentation or control actions, e.g. interlocks, alarms, measurements, trips or redundancy. On the other hand, a HazOp session is a way to enhance the information exchange between the automation designers, process designers and operational personnel	HazOp is often a time-consuming and thus expensive method owing to the systematic procedure of the method and the number of people involved. HazOp is most often applied to continuous processes or process situations. Only minor emphasis is given to the non-continuous processes and process situations like start-up, maintenance.
Action error analysis (AEA)	AEA gives the automation designer proposals for instrumentation to help the personnel to monitor the process during the operation sequences and to prevent the operators from making mistakes.	AEA is an analysis of the technical system, and does not analyse the operator behaviour. The thoughts and intents of the human being, i.e. the reasons for mistakes, are not examined.
Fault tree analysis (FTA)	In FTA, the needs for control or protective actions to diminish the risk can be seen, and the effects of different control and protective actions on safety can be quantified (Hill, 1988). FTA also makes it possible to verify those requirements, which are expressed as quantitative risk values. An advantage of FTA is that it can handle multiple failures. Software fault tree analysis (SFTA) attempts to verify that the program will not in any environment allow a particular unsafe output to occur.	The major risks have to be well known before FTA, for instance, on the basis of a PHA, and specialists are needed for the quantitative analyses. A basic assumption of FTA is that all failures in a system are binary in nature. That is, a component or human being either performs successfully or fails completely. Similarly, the failures are assumed to be instantaneous. In addition, the system is assumed to be capable of performing its task if all sub-components are working.†
Event tree analysis (ETA)	The ETA makes it possible to analyse systems where the chronology of the events is stable and the events are independent. For instance, protective actions performed by automation are often organized in such a chronology and are independent of each other.	A limitation of ETA is that the model consists of the intended actions, and the effects of the failures to conclude the intended actions are modelled. No attention is paid to the possible extra actions, or incomplete actions, including those taken too early or too late.
Failure mode and effect analysis (FMEA)	FMEA studies the potential failures of automation system components and their effects on the function of system. It is also possible to ensure that potential failures will be taken into account, and countermeasures designed for the potential process disturbances that they would cause. An advantage of FMEA is that it is simple and easy to learn.	Like most of the safety analysis methods, FMEA does not study multiple, simultaneous failures without tremendous increase of required labour for studying all the different failure combinations. When failures in automation are analysed, there should be a systematic way to take account of all process situations. This might be difficult.†
Reliability assessment	One purpose of reliability assessment is to check, before the system is commissioned, that it is reliable enough. The weak points of the system are identified, and the efforts can be made to improve them or to diminish the consequences of them. Furthermore, the need for spare parts can be forecast on the basis of the failure rate figures. Reliability assessment is used in connection with safety analysis methods to judge the conformance with the quantitative safety requirements.	Quantitative reliability analyses require specialized reliability engineers and are quite laborious. An uncertainty is to find the data, i.e. the failure rate figures, which are the basis of quantitative reliability analyses.†

† FTA, FMEA and reliability assessment are possible to be done both on the controlled system and on the automation system. When analysing the controlled systems, analyses define requirements for the automation; when analysing the automation the requirements are verified.

design, as well as some of the restrictions and deficiencies are listed. The analyses which are covered are hazard and operability study (HazOp), action error analysis (AEA), fault tree analysis (FTA), event tree analysis (ETA), failure mode and effect analysis (FMEA) and reliability assessment.

Different safety analysis methods cover different aspects of automation design. According to accident studies, an important feature in automation safety is the information on process states given to the operator, but safety analysis methods do not consider this aspect explicitly. Only AEA has a comment on it; asking by what way the operator notices his or her mistakes. AEA is an analysis of the technical system, rather than that of operator behaviour. If a more thorough understanding of action errors is needed, other types of studies are required (Rasmussen, 1986, 1990; Rasmussen *et al.*, 1987; Reason, 1990a,b). However, safety analyses provide information on the process disturbances which might lead to unsafe states, and, in this way the analyses can be used as an aid when designing operator interface. Direct use of the results of safety analyses which are made during the design phases for operator support have been studied, e.g. in Suokas *et al.* (1989).

Another major deficiency of the safety analysis methods is that most of them have been developed for studying continuous process functions. An exception is AEA which studies changing process conditions. However, it is limited to those functions, where human interfere is needed. Another exception is a modified HazOp which is introduced for the non-continuous process functions (Anon., 1977). However, it is not used very widely.

Most of the safety analysis methods study hardware aspects. Some attempts to apply safety analysis to software exists. Software fault tree analysis (SFTA) is an extension of fault tree where the TOP event is a critical software fault determined, e.g. by the system fault tree; and the software is studied backwards through the program to the software inputs That is, SFTA attempts to prove that the program will not in any environment allow a particular unsafe output to be occurred. (Leveson and Harvey, 1983; Cha *et al.*, 1988). The concept of software reliability differs to some extent from the reliability of hardware. This is mainly because of the different nature of failure or fault. There are a number of views as to how software reliability should be quantified. A commonly used approach is to use an analytical model whose parameters are generally estimated from available data on software failures (Goel, 1985; Dale, 1991; Laprie and Littlewood, 1991). The measure of software reliability is supposed to be useful in planning and controlling resources during the development process so that high quality software can be developed. It is also supposed to be useful as a measure for giving confidence in software correctness. However, it is not used in software production in industries. The concept of software reliability, and the reliability models are also discussed in Rook (1990) and Musa *et al.* (1987).

Most of the safety analyses study the possibility of stochastic failures. Because software failures are always systematic, SFTA and software reliability assessment can be considered to cover systematic failures. However, other methods, like HazOp, FMEA and FTA reveal deficiencies in the system, e.g. needs for redundancy, checks, alarms; and these deficiencies are often some kind of systematic faults.

A drawback of most of the methods is that multiple failures are not studied systematically. Especially, simultaneous feailures of different type, such as a human error in connection with a component failure, are difficult to study with a single method. However, human errors, especially, are more probable during abnormal situations than during the normal course of actions. Methods which include a more formal modelling of the hazards may reveal these multiple failure situations. Such methods for hazard identification and modelling are FTA, ETA and reliability assessment which can, in principle, also include modelling of human mistakes and omissions. However, none of the safety analysis methods mentioned here systematically study human aspects. Aspects which are covered by different safety analysis methods are shown in Table 6.

It is obvious that no safety analysis method covers all aspects of safety design. Each method has its own target and is applicable to specific problems, though in principle, it is possible to find a combination of methods which is optimal for each design problem.

Conclusions

The safety and availability performance of process automation affects the flexibility and profitability of production. Accidents and injuries not only cause economic losses, they may cause human suffering and environmental damage. To judge whether the design is safe enough or should be improved, the designer must have a concept of what kind of safety and availability performance is required.

Automation designers do not get data on the safety requirements for automation from the process engineers in a form which enables direct verification. The automation engineer has to analyse the hazards associated with the process, safety analysis methods can be used here. The hazard and operability study is one of the most common methods applied to analyse the safety of process engineering. Other methods adaptable to automation engineering, and which could be used in co-operation with process engineers,

TABLE 6. THE OUTLINED SAFETY ANALYSIS METHODS COVER DIFFERENT SAFETY RELATED ASPECTS OF AUTOMATION DESIGN

	HazOp	AEA	FTA	ETA	FMEA	Reliability assessment
Identification of potential accidents.	■	■		○	■	
Identification of the causes of accidents.	■	■	■	○	■	
Identification of the consequences of failures.	■	■		■	■	
Hardware failures.	○		■	■	■	■
Software failures.			■			■
Quantitative aspects.			■	■	○	■
Human errors.	○	■	○	○		○
Systematic failures	■	○	○	○	○	
Stochastic failures.	■	○	■	■	■	■
Batch processes.	■	■				

■. Covers systematically.
○: Covers occasionally or implicitly.

are action error analysis, fault tree analysis and event tree analysis. The reliability and safety features of the automation or control system can be further studied by means of failure mode and effect analysis. Quantitative reliability assessment can be combined with other methods to get comparative values for the hazards.

This paper has discussed safety of automation and the use of safety analysis methods in automation design. The safety of computer controlled processes includes the question of software safety. Fault tree analysis and reliability assessment have modifications to analyse software. Even if the other methods which are mentioned here do not study software safety explicitly, they support the application software development process by providing methods for requirement specification. According to accident studies, an important feature in automation safety is the information on process states given to the operator. The safety analysis methods do not consider this aspect explicitly. However, safety analyses provide information on the process disturbances which might lead to unsafe states, and, thus the analyses can be used as a standpoint when designing operator interface. Another major deficiency of the safety analysis methods is that most of them have been developed for studying continuous process functions. Furthermore, a drawback of some of the methods is that multiple failures are not studied systematically. Especially simultaneous failures of different type, e.g. a human error in connection with a component failure, are difficult to study with a single methodology. Only methods which include a more formal modelling of the hazards will reveal these multiple failure situations. However, even if safety analyses do not guarantee that all potential hazards will be revealed and countermeasures planned, they provide a practical way for automation engineers to discuss systematically with process engineers and operators the intended and unintended functions and states of the process. Thus, safety analyses enhance information transfer between process engineers, plant operators and automation engineers.

References

Anon (1977). A guide to hazard and operability studies. London, Chemical Industry Safety and Health Council of the Chemical Industries Association Limited

Anon. (1985). Nomenclature for hazard and risk assessment in the process industries. The Institution of Chemical Engineers.

Anon. (1986). Personsikkerhed ved mikroprocessor styringar, Reports ECR-184, ECR-185. ElektronikCentralen, DK-2970, Horsholm.

Anon. (1987a). Personal safety in microprocessor control systems, Summary. København, Nordisk Ministerråd.

Anon. (1987b). PES Programmable electronic systems in safety related applications, 1. An introductory guide. Health and Safety Executive.

Anon. (1987c). PES Programmable electronic systems in safety related applications, 2. General technical guidelines. Health and Safety Executive.

Anon. (1989). Guidelines for chemical process quantitative risk analysis. Center for Chemical Process Safety of the American Institute of Chemical Engineers. New York.

Bloomfield, R. E. and J. Brazendale (1990). SafeIT, 2. Standard framework.

Cha, S. S., N. G. Leveson and T. J. Shimeall (1988). Safety verification in Murphy using fault tree analysis. *Proc. of the Int. Conf. on Software Engineering*, Singapore, pp. 377–386. IEEE Press, New York.

Dale, C. (1991). The assessment of software reliability. *Reliability Engineering and System Safety*, **34**, 91–103.

Dhillon, B. S. and C. Singh (1981). *Engineering Reliability*. John Wiley, New York.

Goel, A. L. (1985). Software reliabiity models: assumptions, limitations, and applicability. *IEEE Trans. on Software Engineering*, **SE-11**, 1411–1423.

Hill, R. (1988). The design of alarms and interlocks—process aspects. *Loss Prevention in Process Industry*, **1**, 25–30.

IEC 1025 (1990). Fault tree analysis. International Electrotechnical Commission.

IEC 812 (1985). Analysis techniques for system reliability, Procedure for failure mode and effect analysis (FMEA). Standard IEC-812.

IEC (1989a). Software for computers in the application of industrial safety-related systems. Standard Draft SC65A, WG9.

IEC (1989b). Functional safety of programmable electronic systems: Generic aspects. Standard Draft SC65A, WG10.

Laprie, J.-C. and B. Littlewood (1991). Quantitative assessment of safety-critical software: why and how? PSAM Conference, Beverly Hills, CA, U.S.A. 4–7 February 1991.

Kletz, T A. (1983). Human problems with computer control, Hazard prevention. *J. System Safety Society*, 24–26.

Lees, F. P. (1980). *Loss Prevention in the Process Industries*. Vol. 1. Butterworths, London.

Lehtelä, M. (1990). Computer-aided failure mode and effect analysis of electronic circuits. *Microelectron. Reliab.*, **30**, 761–773.

Leveson, N. G. and P. R. Harvey (1983). Analysing software safety. *IEEE Trans. on Software Engineering*, **SE-9**, 569–579.

Leveson, N G. (1984). Software safety in computer-controlled systems. *Computer*, 48–55.

Leveson, N. G. (1986). Software safety: why, what, and how? *Computing Surveys*, **18**, 125–163.

Malasky, S. W. (1982). System safety: technology and application. Garland STPM Press, NY, 2nd ed.

MIL (1984). System Safety Program Requirements. MIL-STD-882B.

MIL 217 (1986). Military handbook 217E. Reliability prediction of electronic equipment. Department of Defense. WA.

Musa, J. D., A. Iannino and K. Okomuto (1987). Software reliability, measurement, prediction, application. McGraw-Hill, Singapore.

Nimmo, I., S. R. Nunns and B. W. Eddershaw (1987). Lessons learned from the failure of a computer system controlling a nylon polymer plant. In B. K. Daniels (Ed.), *Achieving Safety and Reliability with Computer Systems*, pp. 189–206. Elsevier, London.

O'Connor, P. D. T. (1981). *Practical Reliability Engineering*. Heyden, London.

Pipatti, R. (1989). Major hazards in industrial processes. Technical Research Centre of Finland. Research Reports 584, Espoo (In Finnish).

Pitblado, R. M., L. Bellamy and T. Geyer (1989). Safety assessment of computer controlled process plants. *6th Int. Symposium Loss Prevention and Safety Promotion in the Process Industries*, 19–22 June 1989 pp. 46, 1–14.

Rasmussen, B. (1986). Unwanted chemical reactions in the chemical process industry. A study of accident case stories. Risø National Laboratory, Internal report.

Rasmussen, J. (1986). Information processing and human-machine interaction. *An Approach to Cognitive Engineering*. Elsevier, London.

Rasmussen, J., K. Duncan and J. Leplat (Eds) (1987). *New Technology and Human Error*. John Wiley, New York.

Rasmussen, J. (1990). The role of error in organizing behaviour. *Ergonomics*, **33**, 1185–1199.

Reason, J. (1990a). Managing the management risk: new approaches to organizational safety. Paper for Workshop on Managing New Technologies, Bad Homburg, May 3–5.

Reason, J. (1900b). *Human Error*. Cambridge University, New York.

Rook, P. (Ed.) (1990). *Software Reliability Handbook*. Elsevier, London.

Rosenof, H. P. and A. Ghosh (1987). Batch process automation. *Theory and Practice*. Van Nostrand, New York.

Shen, K.-C. (1986). On the exploratory study of reliability and safety engineering techniques and their implementation in the machine design process. Institut för maskinkonstruktion, Lunds Tekniska Hǧskola, Lund.

Suokas, J. (1988). The role of safety analysis in accident prevention. *Accid. Anal. and Prev.*, **20**, 67–85.

Suokas, J., R. Fjellheim and R. Wennersten (1989).

Knowledge based risk management system for process industry. *Proc. of the Second Scandinavian Conf. on Artificial Intelligence* SCAI '89, Tampere, 13–15 June.

Taylor, J. R. (1981). Analysis of human error in process plant operation and maintenance. Seminar in human factors and risk analysis. Society of reliability engineers, Scand. Chapter. Stockholm, 14–15 December.

Toola, A. (1988). Reliability analysis of the control system of a conventional power plant. *Reliability Engineering and System Safety*, **21**, 77–89.

Watson, I. A. (1989). Safety and reliability procedures in various industries. Safety and Reliability Directorate, U.K. Atomic Energy Authority, Warrington.

Wells, G. L. (1980). *Safety in Process Plant Design*. George Godwin Ltd. London.

Automatica, Vol. 29, No. 2, pp. 549–554, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Technical Communique

Controller Design Using Fuzzy Logic—A Case Study*

KARL HEINZ KIENITZ†

Key Words—Control systems; optimization; fuzzy logic.

Abstract—Controller design is considered for system specifications which are not handled naturally by analytical methods. Using fuzzy sets and related theory, system specifications are translated into preference functions which are readily combined with search methods to determine adequate controller parameters. This contribution integrates the discussion of the theory and its step-by-step application to aircraft control during the flare-out phase of landing.

1. *Introduction*

IN MANY ENGINEERING applications controller design, tuning and adjustment have been necessary to guarantee the attainment of a variety of application specific performance and stability requirements.

The control engineer, in general, is able to formulate control specifications informally, using some sentences and, eventually, some sketches. Nevertheless a methodic approach to the design, adjustment or tuning problem itself may be intricate, depending on these specifications. As long as quadratic performance indices or eigenvalue specifications are used, for example, controller design is almost straightforward. Often, however, such specifications do not arise naturally and have to be "designed" by themselves to be hopefully equivalent to the original specifications.

In this context it is highly desirable to have a tool for generating a preference function which translates the engineers' "informal" specifications into mathematical language. Such preference function should allow for quantitative control system quality evaluation and be readily combinable with search methods to generate an adequately tuned or adjusted controller.

The main concern of this contribution is to show how fuzzy sets and the related theory may be used to translate relatively informal controller design specifications into a preference function, which is then used to determine adequate controller parameter values using an optimization algorithm such as that due to Nelder–Mead (see references and comments by Himmelblau, 1972). Thus fuzzy sets and the related theory are used as a design decision tool. In the fuzzy set literature several decision applications were reported, such as those by Zimmermann *et al.* (1984). However, a specific application to control engineering is not known to this author.

For better comprehension and illustration of the simplicity and potential of the proposal, this contribution integrates the discussion of the theory and its step-by-step application to a well-known, non-trivial aircraft landing system design problem (Ellert and Merriam, 1963).

The contribution is set up as follows. In Section 2 the case study problem is described. Section 3 presents the chosen controller structure. This choice defines the number of controller parameters to be determined. Section 4 is devoted to the selection of preference functions which truly incorporate the control specifications. In Section 5 the preference function optimization problem over the controller parameter space is treated. Preference function optimization yields adequate parameter values which will satisfy the specifications. Results for the case study problem and final comments are found in Sections 6 and 7, respectively.

2. *Aircraft landing problem definition*

This section closely follows Ellert and Merriam (1963) and Tou (1964).

This case study is concerned with the final phase of aircraft landing, also called the flare-out phase. In this phase the aircraft must be guided along the desired flare-path from a given altitude until it touches the runway. It is assumed that the aircraft is guided to the proper initial conditions for flare-out begin, which in this example range from 80 to 120 ft altitude, and 16–24 ft sec^{-1} descent rate. Variation of descent rate is zero at flare-out begin. The aircraft is waved off for values outside these ranges. At flare-out begin in any case the angle of attack value is supposed to be 12.6°, which is 70% of its stall value. Through adequate control (not discussed herein) aircraft velocity v is maintained constant equal to 256 ft sec^{-1} during flare-out. Only longitudinal motion is considered.

The linearized state equations for the initial angle of attack, adopted henceforth as the nominal value, read:

$$\mathrm{d}x/\mathrm{d}t = Ax + Bu, \quad y = Cx, \quad \begin{bmatrix} \mathrm{d}\theta/\mathrm{d}t \\ \Delta\theta \\ \mathrm{d}h/\mathrm{d}t \\ h \end{bmatrix}$$

where x is the state vector, y is the output and measurement vector, $\mathrm{d}\theta/\mathrm{d}t$ is the pitch rate (rad sec^{-1}), $\Delta\theta$ is variation in pitch relatively to the nominal value (rad), $\mathrm{d}h/\mathrm{d}t$ is rate of ascent (ft sec^{-1}) and h is the altitude (ft). System input u is the elevator deflection (rad). System matrices are given as:

$$A = \begin{bmatrix} -0.6 & -0.76 & 2.96875 \times 10^{-3} & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 102.4 & -0.4 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} -2.375 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10^{-3} & 0 \\ 0 & 0 & 0 & 10^{-3} \end{bmatrix}$$

In addition to the state variables, the angle of attack α is of primary importance since for $\alpha = 18°$ the aircraft reaches stall conditions its geometric relation to the pitch angle and the rate of ascent is well known (see Etkin, 1982):

$$\alpha = \theta - \sin^{-1} \mathrm{d}h/\mathrm{d}t$$

* Received 25 August 1991; recommended for publication by Editor W. S. Levine.

† Department of Control, Instituto Tecnológico de Aeronáutica—ITA, Centro Técnico Aeroespacial—CTA, 12228-900 São José dos Campos SP, Brazil.

The following specifications and control value limitations are given for the flare-out phase of landing.

(a) For reasons of safety and passenger comfort, the desired altitude $h_d(t)$ of the aircraft during flare-out is

$$h_d(t) = \begin{cases} 100e^{-t/5} & 0 \le t \le 15 \\ 20 - t & 15 \le t \le 20 \end{cases}.$$

Thus the desired duration of the flare-out is 20 sec, including 5 sec over the runway.

(b) As a consequence of (a) the desired rate of ascent is the time derivative of $h_d(t)$.

(c) The aircraft must touch the runway with a slightly negative rate of ascent to ensure proper landing.

(d) During landing the angle of attack must remain below the stall value. The aircraft enters the flare-out phase in equilibrium with an angle of attack 0.7 times this value.

(e) The pitch angle $\theta(T)$ at real touchdown time T (not desired touchdown time!) must lie between 0° and 15° to prevent either the nose or the aircraft tail from touching the runway first.

(f) The elevator, which is the only actuator for this problem, has its mechanical stops at −35° and 15°.

3. *Selecting the control structure*

To satisfy requisites (a)–(e) under restriction (f), all the foregoing section, a linear state feedback controller structure is adopted. The choice of the controller structure depends on control engineering considerations which primarily involve realization simplicity and measure availability. In the case of this aircraft landing problem, measures of all state variables are available, which suggests the use of static state feedback. The use of two mutually reinforcing references $h_d(t)$ and dh_d/dt is natural, since the controlled system is expected to track them. Thus the control function becomes:

$$u = -K_1x_1 - K_2x_2 - K_3(x_3 - dh_d/dt) - K_4(x_4 - h_d),$$

where K_1, K_2, K_3 and K_4 are the controller parameters. In the light of these considerations, the controller design process reduces to finding proper values for these four parameters.

At this point it should be noticed that the design specifications will never be satisfied exactly because of the most likely existence of non-nominal permissible initial conditions of altitude and rate of ascent. Hence the formulation of the problem is uncertain in nature and fuzzy set theory will provide the best means of formulating mathematically which controller performs "best" in the most adequate sense.

4. *Generating a preference function*

Design specifications were formulated in a quite complete fashion in Section 2. In this section the specifications are translated into a preference function, a function that allows for the ranking of sets of parameter values as well as for the determination of the acceptability or not of a parameter value set.

For the sake of generality it is assumed that the desired properties of a successfully controlled system are formulated linguistically in terms of n linguistic variables $X_1, \ldots, X_n$, (see Kandel, 1986) as follows

$$\begin{aligned} &[(X_1 \text{ is } A_{11}) \text{ AND } (X_2 \text{ is } A_{21}) \text{ AND} \ldots \text{AND } (X_n \text{ is } A_{n1})] \text{ OR} \\ &[(X_1 \text{ is } A_{12}) \text{ AND } (X_2 \text{ is } A_{22}) \text{ AND} \ldots \text{AND } (X_n \text{ is } A_{n2})] \text{ OR} \\ &[(X_1 \text{ is } A_{1m}) \text{ AND } (X_2 \text{ is } A_{2m}) \text{ AND} \ldots \text{AND} \end{aligned} \quad (1)$$

The A_{ij} are linguistic values characterized as fuzzy sets on the support set of X_i. For the sample problem a description of desirable landing behavior could read (the choice is not unique):

(*Pitch angle at touchdown* is *positive less than* 15°)
AND (*Maximum angle of attack* is *below stall value*)
AND (*Error in h*(t) is *low*)
AND (*Error in* dh/dt is *low*)
AND (*Time at touchdown* is *about* 20 *sec*)
AND (*Rate of ascent at touchdown* is *slightly negative*)
AND (*Maximum rate of ascent* is *negative or about zero*)

How was this sentence generated from specifications (a)–(e)? As a general characterization of an acceptable landing performance in the light of the specifications. "*Pitch angle at touchdown* is *positive less than* 15°" and "*Maximum angle of attack* is *below stall value*" are specifications (d) and (e), respectively. The terms "*Error in h*(t) is *low*" and "*Error in* dh/dt is *low*" account for the requirements that altitude and rate of ascent should remain reasonably close to those specified in (a) and (b). "*Time at touchdown* is *about* 20 *seconds*" enforces that touchdown should really occur near the time specified in (a). This is necessary to hinder the aircraft either from touching the ground before reaching the runway or from overrunning it. "*Rate of ascent at touchdown* is *slightly negative*" is specification (c). "*Maximum rate of ascent* is *negative or about zero*" simply points out that, in order to correct initial conditions and follow the flare-path, the aircraft should not ascend in a perceptible manner. The reason for this is passenger comfort, which was already implicitly stated in the formulation of the desired flare-path $h_d(t)$. Restriction (f) on elevator excursion is not included in the description of landing behavior since it is supposed that the elevator deflection commanded by the controller will not reach saturation values. This supposition is confirmed later in Section 6.

In the next step the linguistic formulation (1) of the desirable situation has to be translated into a preference function. As mentioned before, a preference function allows for the ranking of different sets of controller parameters; the set which produces system behavior with the highest preference function value being regarded as the most satisfactory one. Thus the membership function of the set of desirable (i.e. satisfactory) controlled systems would lend itself as preference function. To derive this membership function, recall that in the specific case of control systems, the linguistic values taken by X_i in (1) generally will be singletons, which means that only one point x_i of the support set will have membership 1 whereas at all other points membership will assume its minimum value. Hence the membership of a certain controlled system in the set of desirable (i.e satisfactory) systems is the membership function induced by (1), which, according to general definitions of fuzzy logic (Kandel, 1986), is:

$$\mu(x_1, \ldots, x_n) = \max_j \min_i [\mu_{A_{ij}}(x_i)], \quad (2)$$

where x_i is the actual scalar value on the support set of X_i.

There obviously remains the choice of the fuzzy sets which characterize the A_{ij}. For computational efficiency the membership functions of these sets should be chosen as simple as possible, for example as piecewise continuous functions.

In the case of the airplane landing system the X_i in (1) are defined as follows:

$$\textit{Time of touchdown} = T \text{ such that } h(T) = 0$$

$$\textit{Error in } h(t) = \max_{t \in [0, T]} |h_d(t) - h(t)|$$

$$\textit{Error in } dh/dt = \max_{t \in [0, T]} |dh_d/dt - dh/dt|$$

$$\textit{Descent rate at touchdown} = (dh/dt)(T)$$

$$\textit{Maximum rate of ascent} = \max_{t \in [0, T]} dh/dt$$

$$\textit{Pitch angle at touchdown} = \theta(T)$$

$$\textit{Maximum angle of attack} = \max_{t \in [0, T]} \alpha(t)$$

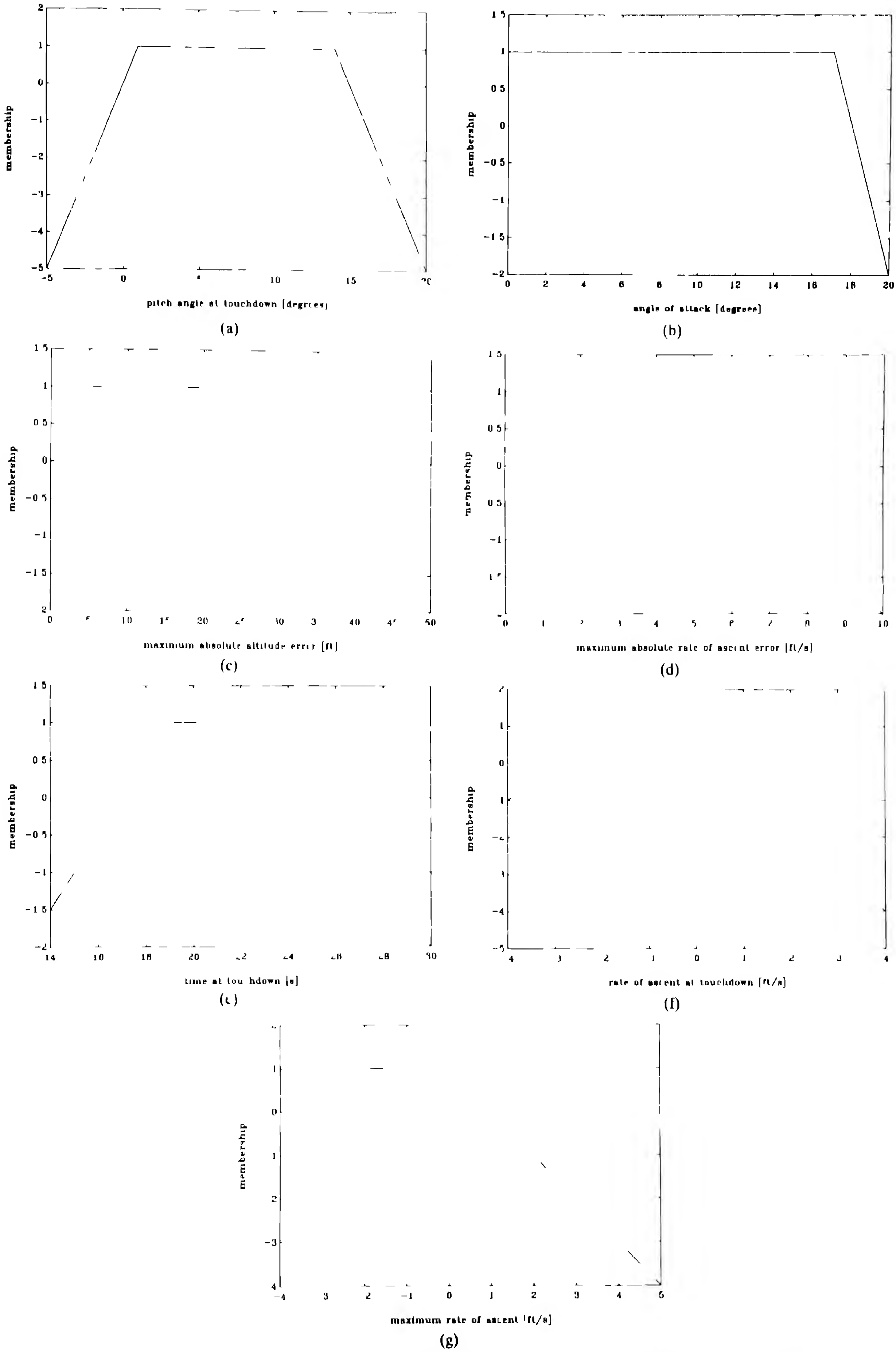


FIG 1 Membership functions for the aircraft landing problem (a) Angles *positive less than* 15° (b) Angles of attack *below stall value* (c) *Low* altitude error (d) *Low* rate of ascent error (e) Touchdown at *about* 20 *sec* (f) *Slightly negative* rates of ascent at touchdown (g) *Negative or about zero* maximum rates of ascent

TABLE 1 INITIAL CONDITIONS

$d\theta/dt$	$\Delta\theta$	dh/dt	h
0	−0 0625	−16	120
0	−0 0781	−20	100
0	−0 0938	−24	80

TABLE 2 SUMMARY OF RESULTS

Order No	Initial guess	Determined parameters	Preference function	Number of iterations
1	0	−0 4311	0 2837	34
	−1	−1 0304		
	−2	−2 4636		
	−1	−1 2422		
2	0	−0 1171	0 1519	25
	−0 5	−0 6675		
	−1 5	−1 7064		
	−1	−0 7630		
3	−0 7158	−1 0087	0 3333	28
	−0 7663	−1 3194		
	−3 5855	−7 6266		
	−0 8224	−1 8916		

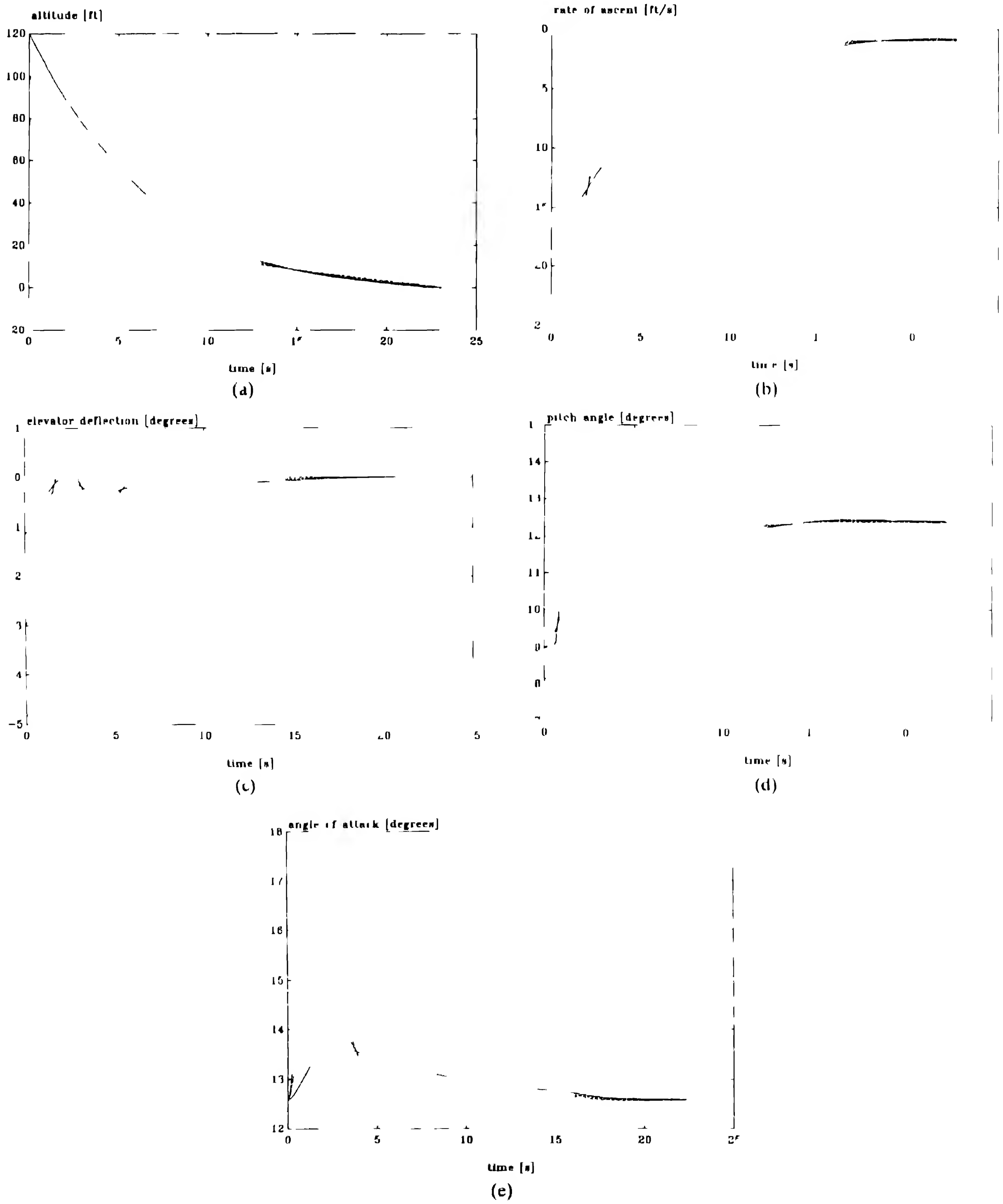


FIG 2 Landing performance for parameter set 2

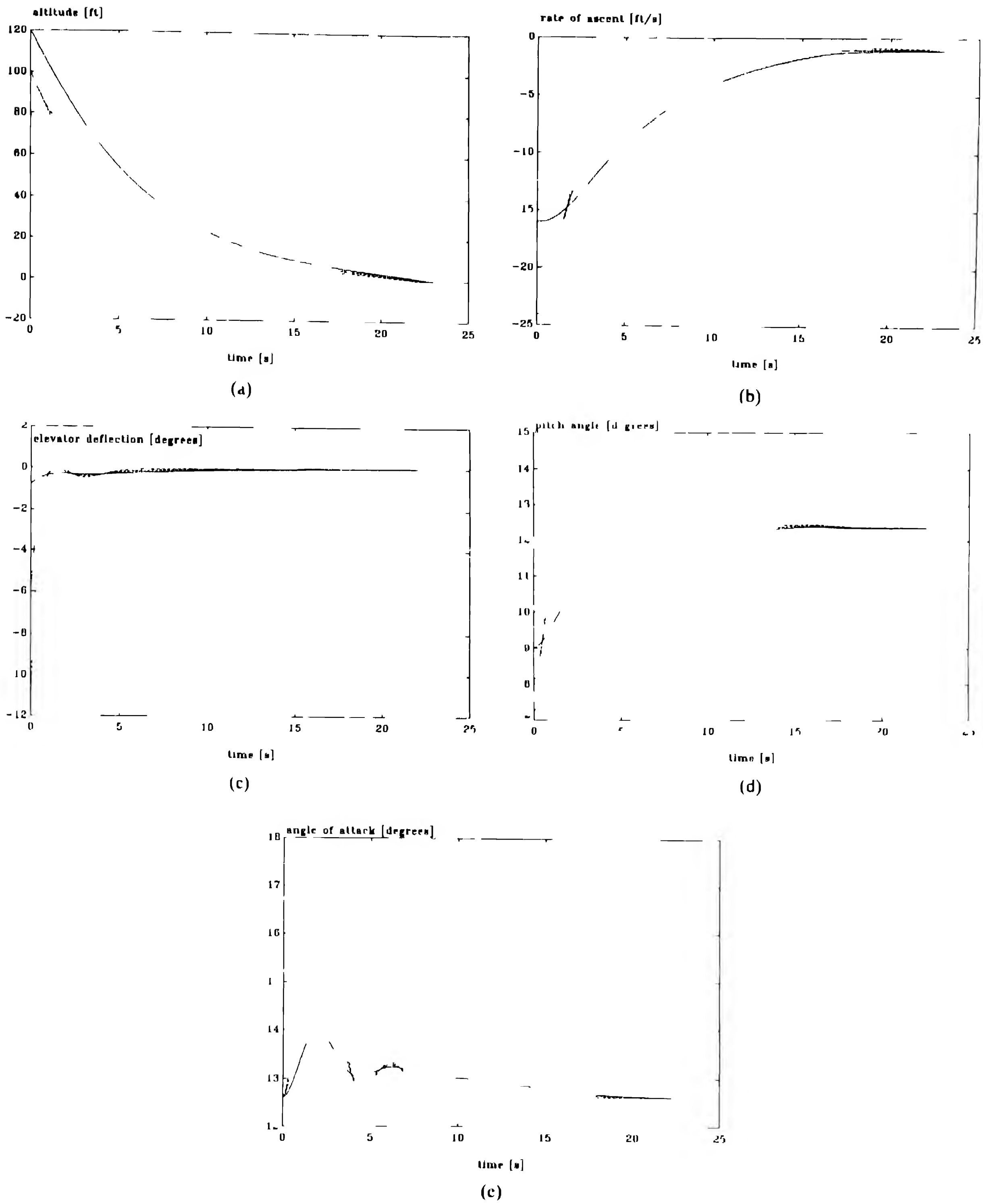


FIG. 3 Landing performance for parameter set 3

The membership functions for the A_i (*low, slightly negative,* etc) are given in Fig 1 Membership value 1 corresponds to fully satisfactory performance with respect to X_i Negative membership characterizes completely unacceptable situations Membership in the range [0, 1] is therefore acceptable and membership 1 is ideal Even the casual reader is generally used to membership values in the interval [0, 1] only However, it is easily seen that for our purpose the exact membership function image is not relevant, as long as it clearly expresses realistic membership ranking Those who dislike negative membership values may map the proposed function onto the interval [0, 1] and continue thereafter to achieve the same results, with an obvious increase in computational expense

For the landing problem, controlled system specifications are given independently of initial conditions, which may vary within the given ranges But system behavior strongly depends on the initial conditions Thus for performance evaluation purposes the nominal and the two worst case initial conditions are considered The preference function is evaluated for those three cases, and the worst value (the lowest) then taken This procedure is adequate since the two

worst case initial conditions are easily identified as those corresponding to the situation in which the aircraft is 20% above or below the desired flare-path at initial time and diverges at the largest permissible rate. The three initial conditions are given in Table 1.

5. *Comments on the optimization process*

The controller selection procedure consists of finding a set of controller parameters which gives a good, possibly the best, preference function value. Starting at an initial guess, a methodic optimization procedure is used to reach a local optimum. Such optimization procedure should not rely on gradient values. The chosen algorithm should be started at several different initial guesses to ensure that a good value of the performance function is reached.

The Nelder–Mead procedure was adopted to determine the controller parameters for the aircraft landing system. At each step system performance was obtained through simulation for the three initial conditions of Table 1. Based on these results, the preference function was then evaluated as described in Section 4.

6. *Results*

The Nelder–Mead algorithm was started at three initial guesses. These initial guesses were determined from root-locus and pole-placement considerations in order to ensure at least system stability. For all three guesses the algorithm located satisfactory controller parameters. The initial guesses, determined parameters, locally optimal values of the preference function and the number of iterations needed by the algorithm to reach the solution are found in Table 2.

Figures 2 and 3 show landing performance for the initial conditions of Table 1 and the calculated parameter sets 2 and 3 of Table 2. Solid curves depict landing performance for the first initial condition of Table 1, dashed curves depict it for the second and dot–dashed curves for the third. Dotted curves are desired performance. Although the system performances in Figs 2 and 3 are similar, the controller gains differ considerably, which clearly influences elevator deflection. Nevertheless it is seen that restriction (f) (Section 2) is not critical.

As indicated by the performance function value, parameter set 3 really is the best since it keeps the aircraft the furthest from stall.

It should be noted that the solution for the landing problem in this contribution is considerably simpler than that determined by Ellert and Merriam (1963) and Tou (1964) and yields better (qualitative) performance.

7. *Conclusions*

It was shown how fuzzy sets and related concepts may help to solve controller design problems via optimization. The illustrated approach is particularly useful when design specifications do not relate directly to quadratic performance indices, system eigenvalues, singular values or other system parameters manipulated by analytic design methods.

If in a particular application certainty factors are involved and/or the X_i in (1), (Section 4) take values which are not singletons, the approach of this contribution needs refinements and extension using the concept of plausible approximate reasoning (see Kienitz, 1990).

References

Ellert, F. J. and C. W. Merriam (1963). Synthesis of feedback controls using optimization theory—an example. *IEEE Trans. Aut. Control*, **AC-8**, 89–103.

Etkin, B. (1982). *Dynamics of Flight Stability and Control*, 2nd ed. John Wiley, New York.

Himmelblau, D. M. (1972). *Applied Nonlinear Programming*. McGraw-Hill, New York.

Kandel, A. (1986). *Fuzzy Mathematical Techniques with Applications*. Addison-Wesley, Reading, MA.

Kienitz, K. H. (1990). Plausible approximate reasoning. *Cybernetics and Systems*, **21**, 647–654.

Tou, J. (1964). *Modern Control Theory*. McGraw-Hill, New York.

Zimmermann, H.-J., L. A. Zadeh and B. R. Gaines (Eds) (1984). *Fuzzy Sets and Decision Analysis*. North Holland, Amsterdam.

Automatica, Vol. 29, No. 2, pp. 555–557, 1993
Printed in Great Britain

0005-1098/93 $6.00 + 0.00

Technical Communique

A Non-conservative Stability Test for 2×2 MIMO Linear Systems Under Decentralized Control*

BORIS E. DAINSON† and DANIEL R. LEWIN‡

Key Words—Multivariable control; decentralized control; PID tuning; stability.

Abstract—This note describes an interaction measure which enables the design of decentralized control for 2×2 linear systems. A sufficient condition is presented which accounts for phase information and therefore provides a less conservative approach than those based on the Small Gain Theorem. A method relying on this new measure permits tighter, independent design of the two controllers, which also allows performance trade-off between the two loops to be accounted for.

1. *Introduction*

DECENTRALIZED CONTROL is probably the most commonly implemented scheme in MIMO systems, because of its relative simplicity and because it is potentially robust to sensor and control actuator failure. The synthesis of such a scheme involves the solution of two problems: (a) the selection of control pairing; (b) the tuning of the feedback controllers. Conditions for *Decentralized Integral Controllability* can be utilized to select adequate pairing (Morari and Zafiriou, 1989). The issue of controller tuning recognizes that any interactions present in the process may cause the performance of the individual control loops to deteriorate. A sufficient criterion for the system stability can be based on the Rijnsdorp Interaction Measure (RIM) (Rijnsdorp, 1965). This convenient method permits the independent tuning of the controllers, where the tuning parameters are constrained by a scalar function whose magnitude depends on the relative level of interaction in the process. One of the sources of conservativeness of the RIM test is due to the fact that it takes into an account only magnitudes of the system elements. Some other interaction measures are described and compared by Grosdidier and Morari (1986). Most of them also use only magnitudes of the matrix elements. For 2×2 control systems we can make the stability test tighter by adding phase information.

2. *Description of the method*

The decentralized control structure illustrated in Fig. 1 can be transformed to the SISO closed loop system given in Fig. 2, by simple block diagram manipulation, where:

$$h_i(s) = \frac{p_{ii}(s)c_i(s)}{1 + p_{ii}(s)c_i(s)}, \quad i = 1, 2, \tag{1}$$

are the closed loop transfer functions of the corresponding SISO loops and

$$\kappa(s) = \frac{p_{12}(s)p_{21}(s)}{p_{11}(s)p_{22}(s)}$$

* Received 20 September 1991; revised 4 February 1992; recommended for publication by Editor W. S. Levine.
† Department of Chemical Engineering, Technion I.I.T., Haifa 3200, Israel.
‡ Author to whom correspondence should be addressed.

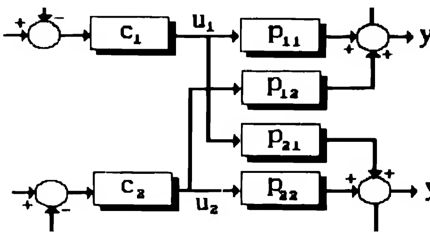


FIG. 1. The decentralized control structure for a 2×2 system.

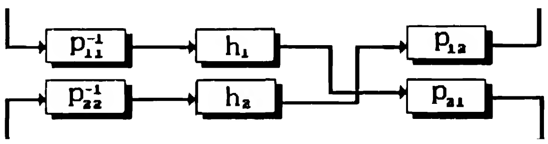


FIG. 2. Transformation of the structure in Fig. 1.

Applying the Nyquist test to this problem, we note that stability can be guaranteed for the positive feedback implied in Fig. 2, so long as the total openloop transfer function does not encircle the (1, 0) point in the complex plane as frequency goes from zero to infinity. Thus, a necessary and sufficient condition of the system's stability can be stated as follows. If $\kappa(s)$ and both of the closed loops $h_i(s)$ are stable, the closed loop MIMO control system is stable if and only if the plot

$$W(i\omega) = h_2(i\omega)\kappa(i\omega)h_1(i\omega), \tag{2}$$

does not encircle (1, 0) point on the complex plane. Unfortunately, this test does not enable independent tuning of the controllers because $W(\omega)$ depends on both of the controllers $c_1(s)$ and $c_2(s)$. It is possible to make the criterion less tight (only sufficient) but comprising of two independent parts, each of which depends only on one of the controllers.

Theorem 1. The Nyquist stability test (2) can be transformed into three simultaneous necessary and sufficient conditions. Closed loop stability is guaranteed if the phase of $W(i\omega)$, $\phi(W)$, is bounded by.

$$0 < \phi(W(i\omega)) < 2\pi, \quad \text{for} \quad 0 \le \omega \le \omega_1^*, \tag{3}$$

the magnitude of $W(i\omega)$ is bounded by:

$$|W(i\omega)| < 1, \quad \text{for} \quad \omega > \omega_2^*, \tag{4}$$

and the critical frequencies are such that: $\omega_2^* \le \omega_1^*$. (5)

Proof. Condition (3) eliminates the possible intersection of the arc of the Nyquist plot with the positive real axis for $0 \le \omega \le \omega_1^*$, while condition (4) ensures no encirclement of the $(1, 0)$ point for $\omega > \omega_2^*$. Thus, condition (5) implies that there is no encirclement of the $(1, 0)$ point for any frequency. Note, that if we neglect all phase information ($\omega_1^* = 0$), the new theorem reverts to a statement of the Small Gain Theorem. □

Because the total phase of the product of complex variables is the sum of their phases, (3) can be estimated as a consequence of the following two inequalities:

$$0 < \phi(\kappa(i\omega)) + \phi(h_1(i\omega)) + \phi(h_2(i\omega)) < 2\pi, \quad \text{for} \quad \omega \le \omega_1^*, \tag{6}$$

or

$$-\phi(\kappa(\omega)) < \phi(h_1(i\omega)) + \phi(h_2(i\omega)) < 2\pi - \phi(\kappa(\omega)), \quad \text{for} \quad \omega \le \omega_1^*. \tag{7}$$

The condition (7) can be transformed into two independent conditions, one for each loop, by introducing a parameter, $\epsilon_1 (0 < \epsilon_1 < 1)$, which defines the trade-off of performance in favor of the first loop. Then, the condition can be rewritten:

$$\left.\begin{aligned} -\epsilon_1\phi(\kappa(\omega)) &< \phi(h_1(i\omega)) < \epsilon_1(2\pi - \phi(\kappa(\omega))) \\ -(1-\epsilon_1)\phi(\kappa(\omega)) &< \phi(h_2(i\omega)) < (1-\epsilon_1)(2\pi - \phi(\kappa(\omega))) \end{aligned}\right\} \quad \text{for} \quad \omega \le \omega_1^*. \tag{8}$$

Similarly, inequality (4) follows from the following condition:

$$\left.\begin{aligned} |h_1(\omega)| &< |\kappa(\omega)|^{-\epsilon_1} \\ |h_2(\omega)| &< |\kappa(\omega)|^{-(1-\epsilon_1)} \end{aligned}\right\} \quad \text{for} \quad \omega > \omega_2^*. \tag{9}$$

Each of the inequalities (8) and (9) contain only one of the controllers, therefore they form a sufficient stability test for independent tuning of decentralized controllers. The proposed criterion is less conservative than the RIM test, which is equivalent to condition (9) alone when $\epsilon_1 = 0.5$ and $\omega_1^* = 0$. It is recommended that the value of ϵ_1 be initially selected as $\epsilon_1 = 0.5$, indicating no preference to a particular loop performance. Increasing its value will bias the closed loop performance in favor of the first loop.

3. *Example*

Consider the following 2×2 system, where the time constants are given in minutes:

$$P(s) = \begin{bmatrix} \dfrac{5}{(4s+1)} & \dfrac{2.5e^{-5s}}{(15s+1)(2s+1)} \\ \dfrac{-4e^{-5s}}{(20s+1)} & \dfrac{1}{(3s+1)} \end{bmatrix}. \tag{10}$$

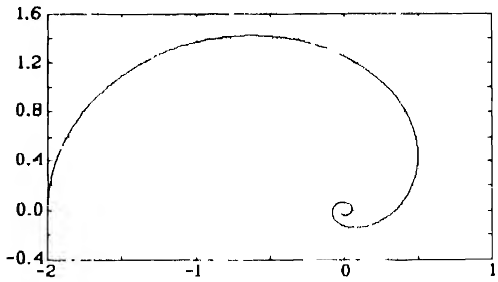


FIG. 4. The Nyquist plot for the example problem with $\lambda_i = 0.1$ min.

In this case, diagonal pairing with integral action cannot be guaranteed stable by the RIM test because $|\kappa(0)| > 1$. The new stability criterion on the other hand, shown graphically in Fig. 3, indicates that closed loop stability can be guaranteed even with arbitrarily high gains. A value of $\epsilon_1 = 0.5$ was selected, indicating no particular preference between the two loops. PI diagonal controller parameters were selected according to IMC tuning rules (Rivera *et al.*, 1986) with the gain chosen to be arbitrarily high (the IMC filter values used where $\lambda_1 = \lambda_2 = 0.1$). As shown in the figure, $\omega_1^* = 0.13\ \text{rad min}^{-1}$ and $\omega_2^* = 0.06\ \text{rad min}^{-1}$, thus satisfying equation (5) and therefore guaranteeing stability. In this case the complementary sensitivity functions of the loops coincide:

$$h_i(s) = \frac{1}{(\lambda_i s + 1)}, \quad i = 1, 2 \tag{11}$$

The Nyquist plot for the diagonally paired system as tuned using the new method, is presented as the contour for $W(i\omega)$ in Fig. 4, showing that the $(1, 0)$ point is not encircled (as predicted).

In fact, it is difficult to use the norm-based RIM test successfully even for the off-diagonal pairing. Shown in Fig. 5, are magnitude and phase Bode plots for the new stability test applied to off-diagonally tuned PID controllers, with the controller gains for the two loops arbitrarily set to $K_{c1} = 0.7$ and $K_{c2} = -0.2$, for which it can be shown that each SISO loop is stable. Again, the PID parameters are IMC-based (Loop 1: reset = 0.086 min^{-1}, rate = 1.08 min, Loop 2. reset = 0.044 min^{-1}, rate = 2.22 min. The controller gains are tunable). The resulting analysis clearly shows that on the basis of the RIM test alone (Fig. 5(a)), nothing can be said about the closed loop stability in the frequency range $0.05 < \omega < 0.35\ \text{rad min}^{-1}$, since here, the sufficient stability condition is violated. On the other hand, the new approach shows that $\omega_1^* > 1\ \text{rad min}^{-1}$, while $\omega_2^* = 0.35\ \text{rad min}^{-1}$, indicating that the system is stable. This result can be confirmed by simulation.

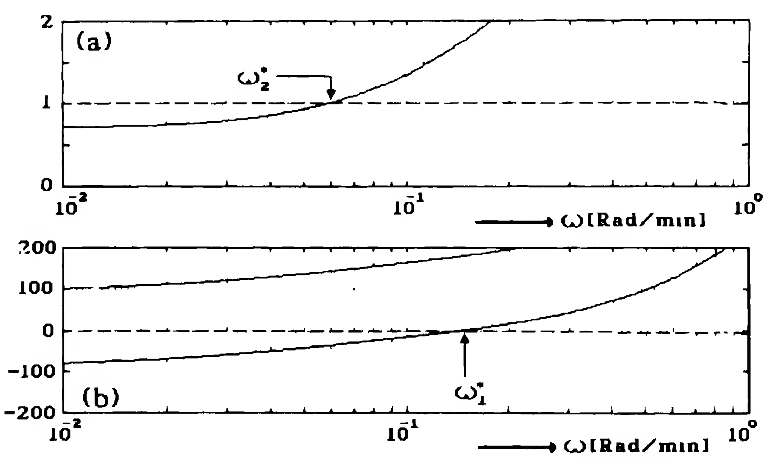


FIG. 3. The new stability test performed on the example problem with $\epsilon_1 = 0.5$. This analysis is for diagonal pairing. Shown are: (a) magnitude, with: $-1/\sqrt{|\kappa|}$ ---- $|h_1|$ and $|h_2|$ (coincident in this case); (b) phase (in degrees): with: — upper and lower bounds (from equation (8)); ---- $\phi(h_1)$ and $\phi(h_2)$, again coincident.

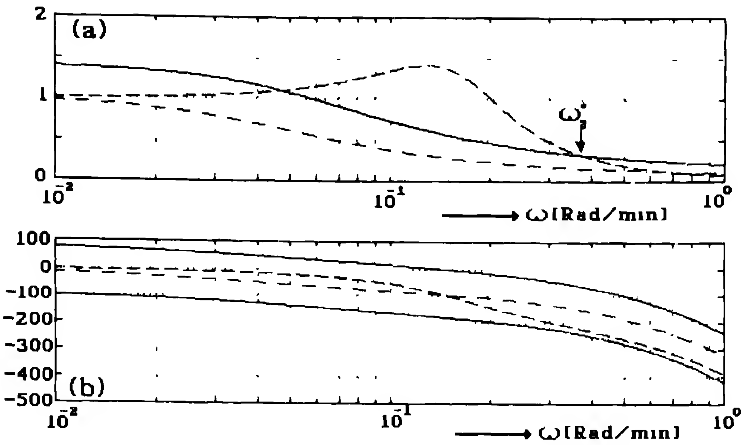


FIG. 5. The new stability test performed on the example problem with $\epsilon_1 = 0.5$. This analysis is for off-diagonal pairing. Shown are: (a) magnitude, with: $-1/\sqrt{|\kappa|}$ $|h_1|$ ---- $|h_2|$; (b) phase (in degrees) Bode plots The upper and lower bounds on the phase plot are those from equation (8).

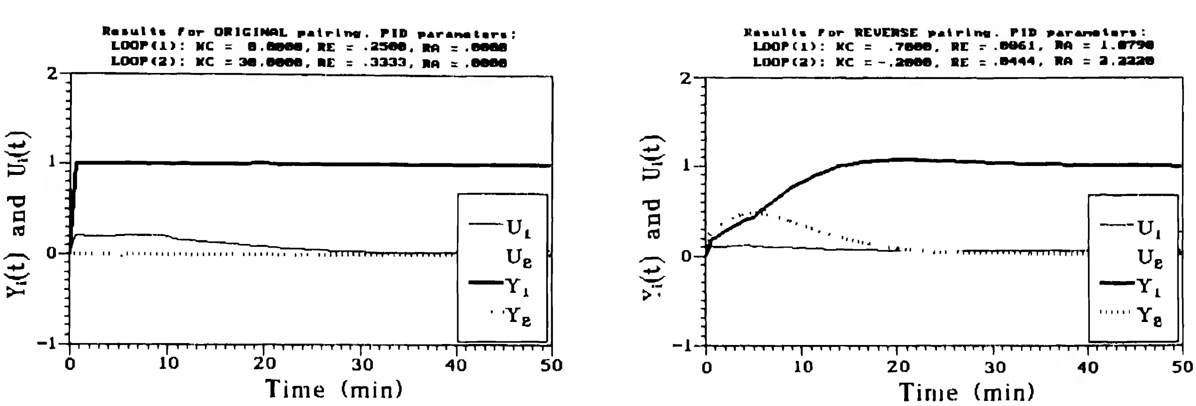


FIG. 6. Comparing the closed loop response for diagonal (left) and off-diagonal pairing (right) Both control schemes were tuned using the new stability criterion with $\epsilon_1 = 0.5$.

Figure 6 compares the closed loop responses achieved using the diagonal pairing (1–1, 2–2) with the off-diagonal (1–2, 2–1) pairing, both tuned as described above The performance of the off-diagonal scheme is severely limited by the significant process delay times, while the diagonal scheme can respond almost instantaneously to set point changes.

4. *Conclusions*

This paper has outlined a proposed sufficient and necessary stability condition which can be used to design decentralized control for a 2×2 MIMO system. The method has been shown to be less conservative than alternative norm-based methods, because unlike them, it conserves phase information. The sufficient and necessary condition derived has been modified to provide a sufficiency condition, still utilizing phase information, which allows the independent tuning of the parameters of the two controllers A trade-off parameter must be set, which delineates the relative importance of the performance of one loop over the other.

References

Grosdidier, P. and M. Morari (1986) Interaction measures for systems under decentralized control, *Automatica*, **22**, 309–316.
Morari, M. and E Zafiriou (1989). *Robust Process Control* Prentice-Hall, Englewood Cliffs, NJ.
Rijnsdorp, J. E. (1965) Interaction in two-variable control systems for distillation columns—I. *Automatica*, **1**, 15–28.
Rivera, D. E., M. Morari and S. Skogestad (1986). Internal model control 4. PID controller design. *I and E. C. Proc. Des. Dev.*, **25**, 252–265.

Automatica, Vol. 29, No. 2, p. 559, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Technical Communique

Comments on 'On Absolute Stability and the Aizerman Conjecture'*

XIONG KAIQI†

Key Words—Absolute stability; Aizerman conjecture; control systems; Lyapunov function.

Abstract—This paper discusses the absolute stability of control system. We point out that Theorem 2 in the paper by Grujic is not correct.

Explanation and correction

IN THEOREM 2 of the paper (Grujic, 1981) the author obtained a necessary and sufficient condition for system (1) to be absolutely stable. This short paper will show that this result is invalid. For this purpose, we are going to give a counterexample.

Consider Example 1 of the paper (Grujic, 1981)

$$\frac{dx}{dt}=\begin{pmatrix}-11, & 0\\ -1 & -10\end{pmatrix}x+\begin{pmatrix}0.4 & 0.8\\ 0.4 & 0.8\end{pmatrix}f(\Sigma),\quad \Sigma=x,$$

$$f(\Sigma)=\begin{pmatrix}\phi_1(\Sigma)\\ \phi_2(\Sigma)\end{pmatrix},\quad \frac{\phi_1(\Sigma)}{\sigma_1}\in[0,2],\quad \frac{\phi_2(\Sigma)}{\sigma_2}\in[0,2],\quad \forall\Sigma\in R^2,$$

and take that Lyapunov function as follows

$$v(x;f)=2|x_1|+|x_2|+20\int_0^{|x_1|}\frac{\phi_1(\pi)}{x_1}\,dx_1+10\int_0^{|x_2|}\frac{\phi_2(x)}{x_2}\,dx_2.$$

In particular, let

$$\phi_1(x)=\begin{cases}2x_1, & x_1\geq 0\\ 0, & x_1<0,\end{cases}\qquad \phi_2(x)=\begin{cases}0, & x_2\geq 0\\ 2x_2, & x_2<0.\end{cases}$$

Then, $v=42|x_1|+|x_2|$, $D^+_{(1)}v^f(x)=42\delta_1(-11x_1+0.4\phi_1+0.8\phi_2)+\delta_2(-x_1-10x_2+0.4\phi_1+0.8\phi_2)$, where

$$\delta_i\triangleq\delta_i(x_i,\dot{x}_i)=\begin{cases}1, & x_i>0 \text{ or } x_i=0 \text{ and } \dot{x}_i>0,\\ 0, & x_i=0 \text{ and } \dot{x}_i=0\\ -1, & x_i<0 \text{ or } x_i=0 \text{ and } \dot{x}_i<0.\end{cases}\quad (i=1,2)$$

For $x_1=-\frac{1}{463}<0$, $x_2=-100<0$, we have

$$\begin{aligned}D^+_{(1)}v^f(x)&=33.6\delta_1\phi_2-10\delta_2x_2+0.8\delta_2\phi_2\\&\quad+462\delta_1(-x_1)+\delta_2(-x_1)\\&=33.6\times200-10\times100+0.8\times200-1=5879>0.\end{aligned}$$

This means that by the set $F_0(L)$ of (7) we cannot ascertain if system (1) is absolutely stable in Theorem 2

A mistake of Theorem 2 results from an error in (A.6) (Grujic, 1981). In fact, (A.6) must be corrected to the following

$$D^+_{(1)}v^f(x)=b^T(I_n+TN(|x|)D^+_{(1)}|x|.$$

But, we can correct Theorem 2 to the following form

Theorem 2*. If C is the identity matrix and L is a compact set then for $F_0(L)$

$$F_0(L)=\left\{v^f: v(x,f)=b^T\left(|x|+T\int_0^x N^*(y)\,dy\right),\quad f\in N_0(L)\right\},$$

where

$$N^*(x)=\operatorname{diag}\left\{\frac{\phi_1(\Sigma)}{\sigma_1\operatorname{sgn}x_1}\ \frac{\phi_2(\Sigma)}{\sigma_2\operatorname{sgn}x_2}\ \cdots\ \frac{\phi_m(\Sigma)}{\sigma_m\operatorname{sgn}x_m}\right\}\Big|_{\Sigma=C}$$

To be a Lyapunov functional family of the system (1) on $N_0(L)$ it is necessary and sufficient that

(i) the vector $[(I_n+LT)b]$ be positive for every $L\in L$, and
(ii) the vector $[R(L,T)b]$ be negative for every $L\in L$.

As is similar to the proof of Theorem 2 in Grujic's paper, we easily prove that Theorem 2* holds.

Reference

Grujic, Lj. T. (1981). On absolute stability and the Aizerman conjecture. *Automatica*, **17**, 335–349.

* Received 25 September 1991; accepted 25 September 1991. Recommended for publication by Editor W. S. Levine.
† Department of Mathematics, Fuzhou University, Fuzhou, Fujian 350002, China.

Automatica, Vol. 29, No. 2, p. 561, 1993
Printed in Great Britain.

0005-1098/93 $6.00 + 0.00

Technical Communique

Reply to "Comments on 'On Absolute Stability and the Aizerman Conjecture'"*

LJUBOMIR T. GRUJIĆ†

Key Words—The Lyapunov method; the Lyapunov functions; Lurie systems; absolute stability; Aizerman conjecture; stability domains.

Abstract—The note presents the refinements of some results of Grujić (1981) and of comments by Kaiqi (1992).

COMMENTS BY Kaiqi (1992) on Grujić (1981) are correct under the following refinments.

Kaiqi (1992) contributes by introducing a new tentative Lyapunov functional family that should be denoted by $F_0^*(L)$ rather than by $F_0(L)$, where

$$F_0^*(L) = \left\{v^f : v(x;f) = b^T\left(|x| + T\int_0^x N^*(y)\,dy\right), f \in N_0(L)\right\},$$

with $N^*(x)$ defined by

$$N^*(x) = \operatorname{diag}\{n_1^*(x_1), n_2^*(x_2) \cdots n_m^*(x_m)\},$$

$$n_i^*(X_i) = \begin{matrix} \phi_i(x_i)\,|x_i|^{-1}, & x_i \neq 0, & \\ 0, & x_i = 0, & i = 1, 2, \ldots, m. \end{matrix}$$

Theorem 2* by Kaiqi (1992) is a new development of Theorem 2 by Grujić (1981). The latter is correct provided $N_0(L)$ in it, as well as in its proof, Corollaries 1 and 2 and Examples 1 and 2, is everywhere replaced by $N_d(L)$, where $N_d(L)$ is the family of all odd $f \in N_0(L)$.

References

Kaiqi, X. (1992). Comments on 'On absolute stability and the Aizerman conjecture'. *Automatica*, **29**, 558.
Grujić, Lj. T. (1981). On absolute stability and the Aizerman conjecture. *Automatica*, **17**, 335–349.

* Received 8 June 1992; accepted 8 June 1992. Recommended for publication by Editor W. S. Levine.
† Ecole Centrale de Lille, BP 48, 59651 Villeneuve d'Ascq-Cedex, France, on leave from the Faculty of Mechanical Engineering, University in Belgrade, P.O. Box 174, 11000 Belgrade, Serbia, Yugoslavia, until 1 October, 1992.

Automatica, Vol. 29, No. 2, pp. 563–564, 1993
Printed in Great Britain.

0005-1098/93 \$6.00 + 0.00

Technical Communique

Relay Auto-tuning in the Presence of Static Load Disturbance*

C. C. HANG,† K. J. ÅSTRÖM‡ and W. K. HO†

Key Words—Disturbance rejection; limit cycles; PID control; process control; relay control.

Abstract—Static load disturbances during the relay tuning experiment introduce errors in the estimates of the ultimate gain and ultimate period. This paper shows how an automatic bias can be used to overcome the problem.

1. *Introduction*

AUTO-TUNING OF PID controllers has recently received much attention in the literature because of its potential applications in reducing system start-up time, and in tightening process control through regular re-tuning (Bristol, 1977; Kraus and Myron, 1984; Higham, 1984; Hess *et al.*, 1987; Radke and Isermann, 1987; Åström and Hägglund, 1988a; Hang *et al.*, 1991). Many commerical products for auto-tuning have appeared in the market since the mid 1980s.

The relay feedback auto-tuner (Åström, 1982; Åström and Hägglund, 1984a, b, 1988a) is one such product. It is based on the automatic measurement of the ultimate gain and ultimate period (the point of the Nyquist curve that first intersects the negative real axis) from which the PID and PI controller parameters are computed. Its greatest merit is that *a priori* information about the process time scale and dynamic structure is not needed. Furthermore, it can automatically produce an excitation signal with frequencies around the ultimate frequency of the process. Therefore it is also suitable for initializing other more sophisticated auto-tuning or adaptive control algorithms (Lundh and Åström, 1992).

An enhancement of the relay feedback auto-tuner is introduced in this paper. It attempts to overcome the problem of static load disturbance during the relay experiment which results in inaccurate estimates of the ultimate gain and ultimate period.

2. *Static load disturbance*

Static load disturbances are common in process control. Their effects during the relay experiment can be illustrated using the following process

$$G_p(s) = \frac{1}{(1+s)^2},$$

and the step-up shown in Fig. 1.

2.1. *Errors in the estimates of ultimate gain and ultimate period.* When the relay is connected in the feedback loop, for most common processes, oscillation with symmetrical positive and negative half-cycles would occur as shown in the first 10 sec in Fig. 2. When a static load disturbance of magnitude less than the amplitude of the relay occurs, the oscillation become asymmetrical (Hang and Åström, 1988). This is shown in the second 10 sec in Fig. 2 where the relay amplitude and the static load are 0.1 and 0.08, respectively. The static load and the relay amplitude are almost equal and the estimated ultimate gain and ultimate period are in error by +14% and +21%, respectively.

2.2. *Error detection and self-correction.* Because of the periodic nature of the signals, the static load disturbance can be determined easily by considering the dc components of the process input and output. The dc component of the input, u_{dc}, to the process is given by

$$u_{dc} = \frac{y_r}{k_p} + \frac{t_1 - t_2}{t_1 + t_2} d + l, \tag{1}$$

where y_r, k_p, d and l are the set-point, process static gain, relay amplitude and static load disturbance, respectively; t_1 and t_2 are the positive and negative relay output intervals, respectively. The dc component of the process output, y_{dc}, is given by

$$y_{dc} = y_r - \frac{1}{t_1 + t_2} \int_{\tau}^{\tau +} e \, dt, \tag{2}$$

where τ is chosen such that the integration is performed over one period of the steady-state oscillation. Since the dc gain of the process is k_p, hence

$$k_p u_{dc} - y_{dc},$$

and

$$k_p\left(\frac{y_r}{k_p} + \frac{t_1 - t_2}{t_1 + t_2} d + l\right) = y_r - \frac{1}{t_1 + t_2} \int_{\tau}^{\tau + t_1 + t_2} e \, dt. \tag{3}$$

Rearranging (3) gives

$$l = -\frac{t_1 - t_2}{t_1 + t_2} d - \frac{1}{k_p(t_1 + t_2)} \int_{\tau}^{\tau + t_1 + t_2} e \, dt. \tag{4}$$

To restore symmetry, the effect of the static load must be cancelled. Therefore a bias, u_b, equals to the negative of the load should be added to the relay output. Thus

$$u_b = \frac{t_1 - t_2}{t_1 + t_2} d + \frac{1}{\hat{k}_p(t_1 + t_2)} \int_{\tau}^{\tau +} e \, dt, \tag{5}$$

where k_p is replaced by its estimate $\hat{k}_p$.

This formula can be used for retuning when the process gain is already known. For initial tuning, a guess of $\hat{k}_p$ has to be made. In process control, a rule of thumb would be to set $\hat{k}_p = 1$ as the process static gain is usually between 0.5 and 2. If $\hat{k}_p \neq k_p$, the bias calculated using (5) would not cancel the static load completely and the oscillation could still be asymmetrical. However, at this stage, the asymmetry could be slight and hence tolerable; if not, a second bias can be applied. With the bias u_b, (3) is modified to

$$k_p\left(\frac{t_1' - t_2'}{t_1' + t_2'} d + l + u_b\right) = -\frac{1}{t_1' + t_2'} \int_{\tau'}^{\tau' + t_1' + t_2'} e \, dt, \tag{6}$$

where t_1' and t_2' from the new positive and negative relay output intervals, respectively. It now follows from (4), (5)

* Received 30 March 1992; received in revised form 16 October 1992. Recommended for publication by Editor William. S. Levine.

† Department of Electrical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511.

‡ Department of Automatic Control, Lund Institute of Technology, Box 118, S-211 00 Lund, Sweden.

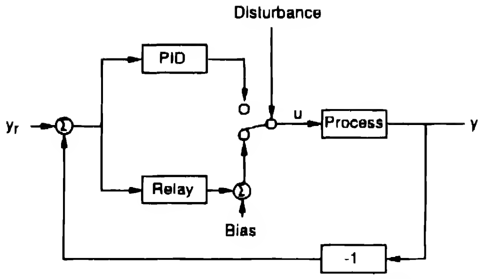


FIG. 1. Set-up for relay auto-tuning.

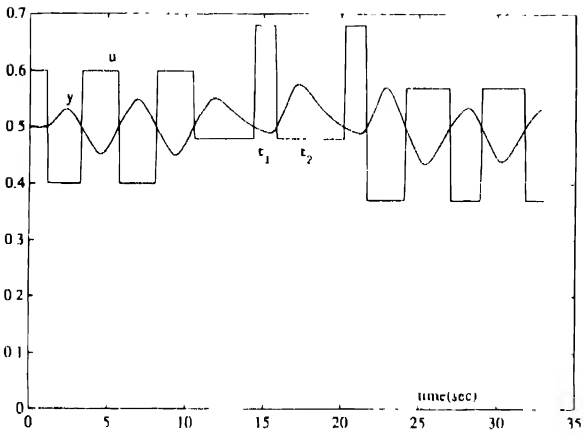


FIG. 2. Static load disturbance and corrective bias with $\hat{k}_p = 0.5$.

and (6) that the process static gain can be obtained from

$$k_p = \frac{(t'_1 + t'_2)\int_T^{T+t_1+t_2} e\,dt - (t_1+t_2)\int_{T'}^{T'+t'_1+t'_2} e\,dt}{(t_1+t_2)(t'_1 - t'_2)d + \frac{1}{\hat{k}_p}(t'_1+t'_2)\int_T^{T+t_1+t_2} e\,dt}. \quad (7)$$

With k_p known, l can be calculated from (4) and a second bias, u'_b can be added to u_b to completely cancel the static load. Hence

$$u'_b + u_b = -l,$$

and

$$u'_b = -(u_b + l).$$

Furthermore, once k_p and l are known, a low order transfer function can be estimated from the wave-form of the oscillation (Åström and Hägglund, 1988b).

The new feature of automatically biasing the relay once the asymmetry is detected can be easily incorporated into the existing relay auto-tuner. Its effectiveness is demonstrated in Fig. 2. At $t = 22$ sec, a bias was applied based on a initial guess of $\hat{k}_p = 0.5$ ($k_p = 1$) which did not result in symmetrical oscillation. However, notice that the asymmetry is slight and the amplitude and period of the oscillation are not too far from those obtained when the oscillation is symmetrical. The estimated ultimate gain and ultimate period are out by only +2% and −2%, respectively. In practice, this could be an acceptable error and the relay experiment stopped at this point. Figure 3 shows for the same system, what could happen given a poor initial guess of $\hat{k}_p$. At $t = 22$ sec a bias was applied based on the initial guess of $\hat{k}_p = 0.25$. It is obvious from the figure that the bias was not effective. At $t = 35$ sec, k_p was estimated using (7) and a second bias was applied. Symmetry in the oscillation was subsequently restored.

The problem of asymmetrical oscillations also occurs for a relay with hysteresis. In this case, all the equations are still valid and the bias computed can likewise be added to the relay to restore symmetry and hence accuracy in the

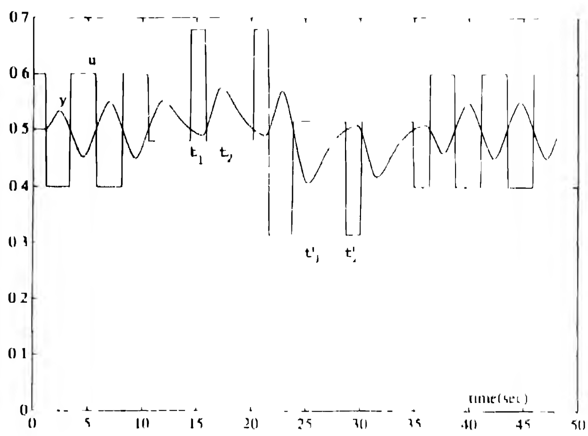


FIG. 3. Static load disturbance and corrective bias with $\hat{k}_p = 0.25$.

estimated parameters. The effects of a sinusoidal load disturbance during the relay experiment was discussed in Hang and Åström (1988). A simple solution has not been found for this problem.

3. *Conclusions*

Auto-tuning of PID controllers using relay feedback is known to be a robust technique which requires little prior knowledge of the process compared to other auto-tuning approaches. To further enhance the existing relay feedback auto-tuner, automatic biasing of the relay to overcome static load disturbance is recommended.

References

Åström, K. J. (1982). Ziegler–Nichols auto-tuners. Report TFRT-3167, Dept. of Automatic Control, Lund Inst. of Tech., Lund, Sweden.

Åström, K. J. and T. Hägglund (1984a). Automatic tuning of simple regulators. *Proc IFAC 9th World Congress*, Budapest, Hungary, pp. 1867–1872.

Åström, K. J. and T. Hägglund (1984b). Automatic tuning of simple regulators with specifications on phase and amplitude margins. *Automatica*, **20**, 645–651.

Åström, K. J. and T. Hägglund (1988a). Automatic tuning of PID controllers. ISA, Research Triangle Park, NC, U.S.A.

Åström, K. J. and T. Hägglund (1988b). A new auto-tuning design. *Proc IFAC Symp. on Adaptive Control of Chemical Processes*, Copenhagen, Denmark, pp. 143–148.

Bristol, E. H. (1977). Pattern recognition: an alternative to parameter identification in adaptive control. *Automatica*, **13**, 197–202.

Hang, C. C. and K. J. Åström (1988). Practical aspects of PID auto-tuners based on relay feedback. *Proc. of IFAC Int Symposium on Adaptive Control of Chemical Processes*, Copenhagen, Denmark, pp. 153–158.

Hang, C. C., K. J. Åström and W. K. Ho (1991). Refinements of the Ziegler–Nichols tuning formula. *IEE Proc. D, Control Theory and Appl.*, **138**, 111–118.

Hess, P., F. Radke and R. Schumann (1987). Industrial applications of a PID self-tuner used for system start-up. *Proc. IFAC World Congress*, Munich, Germany.

Higham, E. H. (1985). A self-tuning controller based on expert systems and artificial intelligence. *Proc. Control 85*, U.K., pp. 110–115.

Kraus, T. W. and T. J. Myron (1984). Self-tuning PID controller uses pattern recognition approach. *Control Engineering*, June, pp. 108–111.

Lundh, M. and K. J. Åström (1992). Initialization of robust adaptive controllers. *Preprints IFAC Symposium on Adaptive Control and Signal Processing*, Grenoble, France.

Radke, F. and R. Isermann (1987). A parameter-adaptive PID controller with stepwise parameter optimization. *Automatica*, **23**, 449–457.

Automatica, Vol. 29, No. 2, pp. 565–571, 1993

Vector Lyapunov Functions and Stability Analysis of Nonlinear Systems*

V. Lakshmikantham, V. M. Matrosov and S. Sivasundaram

Reviewer: P. C. OJHA
Automatic Control Laboratory, University of Gent, Grotesteenweg Noord 2, B-9052 Gent-Zwijnaarde, Belgium.

SINCE THE BEGINNING of the last century mathematicians have sought and with considerable success been able to define physical systems and phenomena using differential equations for exploring the underlying harmony in an apparently diverse nature. Lyapunov's gift to the scientific community, of a tool which is now widely accepted and fruitfully used for the analysis of differential equations, most of which are nonlinear in nature, gave the much required thrust for the development of modern stability theory. While Lyapunov's first method investigates the local stability properties by linearization around a working point, his name is commonly associated with the second, which involves use of continuous functions of state variables. Besides Lyapunov's second method is particularly useful as it does not demand the knowledge of solutions of differential equations but instead measures the domain in terms of distance which is the Euclidean length of the state vector and thus are powerful tools for practical applications.

In this book the authors have sought to collect and present the main trends in the basic theory of the method of Vector Lyapunov functions. The authors begin with the generalization of Lyapunov functions by relaxing the conditions of positive definiteness and decrescentness. Comparison principles are used in order to bring together the concept of Lyapunov functions and the theory of differential inequalities, which in turn enables the transformation of a given complicated differential system to the study of relatively simpler scalar differential equations, thus facilitating the study of stability and other properties of solutions. A brief description of various forms of stability and the ensuing boundedness is given in a very succinct manner and is followed by examples to drive home their relevance. The otherwise theoretical and mathematical treatment of the subject is given a practical touch when the authors set out to define the terms of practically stable systems giving example of a missile which may oscillate around a mathematically unstable course even though its performance is acceptable.

Having set a framework for simplifying the complex nonlinear system the authors then refine the basic Lyapunov theorems using stability results employing two Lyapunov functions thus relaxing the conditions further. Boundedness and practical stability are further proved using two Lyapunov functions, which set the framework for using several Lyapunov functions with a view to generalize the conditions further. This framework is later used to analyse results of global stability due to the fact that the use of several Lyapunov functions basically involves dividing the vicinity of some convenient set into suitable subsets. The authors then drive home the practical relevance of Vector Lyapunov functions by analysing the large scale systems in context of perturbation theory. Perturbed system analysis background is essential to the study of large scale systems since large scale systems constitute of subsystems some of which may be disconnected and reconnected during functioning, thus destroying stability and causing the system to fail. The stability and boundedness properties of perturbed systems are studied using coupled comparison systems. Also the authors show that given good qualitative properties there exists a Lyapunov function for each subsystem and a combination of these Lyapunov functions which constitute the Vector Lyapunov function may then be used to study the whole large composite system. The theory involving perturbed and unperturbed systems is woven together using a new comparison theorem. The restriction of quasimonotone property is taken care of by introducing the concept of quasisolutions, which leads to isolated subsystems thus simplifying analysis and is of immense practical importance.

Having formulated the theory of Vector Lyapunov functions the authors have devoted a considerable part of the book to develop and extend the method of Vector Lyapunov functions to a variety of nonlinear systems. In particular the examples on control systems, where a convenient and unified method of specifying control sets in order to keep the system within a desired practical stability behaviour of the controlled motion, is developed and is relevant to the control community. Also the example on decentralized control system has been treated in the context of perturbation theory. Besides the various examples that are evenly spread over the book, the authors have devoted a chapter to analysing various models of practical relevance. The practical fruitfulness of the method is discussed in wide ranging applications with examples from the economic models based on the Walrasian approach of supply and demand, the motion of aircraft, immunology, chemical kinetics and neural networks. Among the various models discussed in the framework of Vector Lyapunov function are the neural networks and motion of aircraft which are of particular relevance to control theoreticians.

It would be appropriate to add that besides stability analysis the method can be exploited much further by control, adaptive control engineers and theoreticians. Besides global stability analysis a natural application of the theory of Vector Lyapunov function would be to analyse the convergence properties of the system parameters, that is, the extent to which a solution of nonlinear dynamical system may be required to converge to ensure stability and acceptable performance of the adaptive control algorithm. There is little doubt that Lyapunov functions are becoming increasingly popular in the adaptive control community. Also the fact that the theory of Vector Lyapunov function merges so well with neural networks models is an interesting and welcome feature. Since neural networks are being widely used for modelling and identification of input–output spaces of dynamical systems and static identification which include pattern recognition, makes this framework relevant to wide ranging needs. The book though appealing to a wide ranging audience, is written in a mathematical style. While leaving little doubt about the practical and theoretical implications of Lyapunov's second method, a corresponding geometrical or graphical interpretation of theorems would make the reading lucid and within easy grasp of those who are not mathematically inclined.

* *Vector Lyapunov Functions and Stability Analysis of Nonlinear Systems*, Mathematics and its Applications, Vol. 63 by V. Lakshmikantham, V. M. Matrosov and S. Sivasundaram. Kluwer Academic Publications (1991).

The book brings together wide ranging results of considerable relevance. It is particularly useful to theoreticians and theoretically inclined engineers who are looking for tools to analysing the physical systems in a better and unified manner. To sum up the book is a research monograph for those with relevant background in nonlinear systems theory. Nevertheless the book presents the results in a brief, to the point manner and is perhaps a fitting tribute to Lyapunov in a year which marks the century after publication of the original pioneering work of Lyapunov.

About the reviewer

P. C Ojha completed his graduation in engineering in 1988 specializing in electronics and communications from Aligarh University, India where he was an active member in the IEEE student branch. Thereafter he worked as a hardware computer maintenance engineer with a computer manufacturing firm in India. He pursued a masters degree at the Control Systems Centre of University of Manchester Institute of Science and Technology specializing in Control and Adaptive control. He completed his masters in engineering in 1991. At present he is associated with the Automatic Control Laboratory, University of Gent, and is involved with an industrial project. His research interests include identification techniques, nonlinear dynamical systems and neural networks.

Control Sensors and Actuators*

Clarence W. de Silva

Reviewer: M. MACHACEK
APAX Computers Ltd., 10 Chartfield Road, Reigate, Surrey, RH2 7JZ, U.K.

THE STATED author's intention was to summarise the material used in undergraduate and postgraduate courses in control system instrumentation. The book is aimed at both students and practising instrumentation engineers. Enough material was to be assembled in the book from which selections could be made for particular courses.

The sensors and actuators dealt with are generally those used in piece-part manufacturing rather than handling of fluids (process control). Even with this restricted scope, the book provides a wealth of information on over 400 pages of text and figures. Perhaps the title of the book may be a bit misleading. Readers might expect it to cover a wider range of measurements, e.g. temperature, level, flowrate, etc. which it does not.

Sensors and actuators used in robots, machine tools and other devices subject to motion and force control are considered in great detail. Transducers and signal processors are included where necessary. Both analog and digital sensors are described.

The sensors and actuators are always presented as integral parts of control systems, the system integration is an issue in each of the chapters on individual element types.

In Chapter 1, an introduction is given to control system instrumentation. Analog and digital control systems are discussed and the advantages of the digital systems are clearly stated. An assumption is made of the reader's good grasp of control theory. However, the knowledge of the theory is not essential for understanding the bulk of the book—only those parts where integration into control systems is discussed.

An overview of sensor and transducer models is given in Chapter 2. A very good general approach to the flow and effort variables, input/output rating and interface matching is presented and model schemes are set out to which references are made throughout the text. Performance criteria, both in the time and frequency domains are listed. Most of the widely used methods for estimating errors and uncertainties are given in detail. Perhaps some of the flow and effort variables could be explained in more detail in order to illustrate better the principles of analogy between mechanical and electrical variables.

Chapter 3 deals with analog sensors of motion. Most widely used sensors of displacement, velocity and acceleration are described. Jerk is noted as a variable to be measured but no further details are given of how it may be or why it should be measured. The sensors described in detail are those based on variable resistance, reluctance, mutual inductance and capacitance, on permanent magnets, eddy currents and piezoelectricity. Others, noted but not described in detail, are fiber optic sensors, including an optical gyroscope, Doppler interferometers, ultrasonic sensors and mechanical gyroscopes. Proximity sensors are included even though they do not actually provide quantitative measurements

Force and torque analog sensors are dealt with in Chapter 4. Potential conceptual difficulties in determining causality in force control systems are well noted even though a more detailed explanation with additional examples could have been helpful. Good description of strain gauges is given, both resistive and semiconducting, including their transducers. Deflection methods of measuring torque using motion sensors is described and analysed. Tactile sensors are noted as a new technology, with signal processing identified as the main topic of on-going research.

Motion sensing using digital transducers is described in Chapter 5. Incremental and absolute shaft encoders based on optical, electrical, magnetic or proximity sensing are handled in great detail, including their errors. Digital resolvers and tachometers are described. Most of the chapter deals with rotary sensors, linear motion sensors and limit switches are noted.

Permanent magnet and variable reluctance stepper motors are the subject of Chapter 6. Stepping sequences are explained perhaps in too much detail for various geometries of the motor construction—a single general method would have been more in tune with the style of the book. Static and dynamic behaviour analysis and motor models are good enough for understanding of open and feedback loop controls. Damping requirements are analysed, mechanical and electrical damping principles are clearly explained. Good, practical guidance is given for selecting a motor for a given application.

Chapter 7 could easily have been split into several shorter sub-chapters for better clarity. As it is, this chapter deals with the whole range of continuous drives, including electrical DC and AC motors as well as hydraulic valves, motors and pumps. The DC motors, with wired armature and brushless, are handled in greatest detail. Models are given for various field winding connections, issues for open and closed loop control are well explained. Usage of SCRs are well illustrated even though the SCR commutation would have perhaps deserved some more detail. Torque motors, with permanent magnets and geared, are noted. AC induction motors with wound and cage rotors are discussed together with their models and ways of controlling the speed/torque. Of the control methods, the field feedback

* *Control Sensors and Actuators* by Clarence W. de Silva. Prentice Hall, Englewood Cliffs, NJ (1989).

method could have been explained a little better. AC synchronous motors are handled very briefly. Somewhat surprisingly, linear motors and solenoids are included under the heading of AC synchronous motors. Pressure and flow control of hydraulic actuators is handled thoroughly, particulars of the most common elements are noted. Pneumatic controls are noted but not handled in any detail. Guidance for actuator selection is given but only for DC motors—others could have been included for consistency.

The text is full of well-chosen examples that serve to illustrate the theory. There are adequate illustrations to make the text easy to follow and comprehend; in only very few cases one needs to turn a page in order to relate the illustration and its reference in the text. There are some typographical errors of minor nature. The reviewer had no time to examine the numerous equations but some misplaced brackets in the section on statistics were easy to spot.

The great value of the book is in the wealth of problems given at the end of each chapter. The problems generally require a great deal of creative thinking as well as good understanding of the engineering issues. The problems would be of very good value under a tutor's guidance.

References are given at the end of each chapter.

The author has managed to assemble a great deal of material on a wide range of subjects and yet managed to go into sufficient level of detail on all the issues covered. I believe that the aim set out for the book, i.e. to be a collection of material from which individual courses can be assembled, has been well achieved.

Perhaps in some parts the author's expectation of theoretical knowledge of an average practising engineer is a bit high but understanding of all detail is not essential to get benefits from the book. The bulk of the book could be followed by those with a good knowledge of algebra and differential calculus.

About the reviewer

Milos Machacek received his Dipl. Ing. degree in electrical engineering in 1966 from the Czech Technical University of Prague. He worked for four years as a tutor and lecturer at the same university, in the department of electrical machines.

After a move to England in 1970, he worked in various capacities as a research and development engineer, project and development manager for:

- Burroughs Machines Ltd (automatic banking equipment),
- I. H. Williams & Co. (LSI circuit design consultancy),
- Thorn EMI Datatech Ltd (communication equipment),
- Foxboro G.B. Ltd (process control instrumentation and systems).

He now works as independent consultant, specializing in industrial measurements and measurement systems.

Adaptive Filter Theory*

Simon Haykin

Reviewer: BJÖRN WITTENMARK
Lund Institute of Technology, Department of Automatic Control, Box 118, S-221 00 Lund, Sweden

THE BOOK Adaptive Filter Theory gives a very thorough and good treatment of a large and growing area. Adaptive filters are defined as filters where the coefficients of the filters are changing with changing statistics of the signals to be filtered. By using adaptive filters there can be significant improvements compared with conventional filters with fixed characteristics. Adaptive filters have successfully been applied in, for instance, communications, control, radars, sonars, seismology, and biomedical engineering.

The author has a three-fold aim of the book:

(1) develop the theory of different classes of linear finite impulse response (FIR) adaptive filters,
(2) to illustrate the theory with examples and computer experiments,
(3) introduce some ideas of nonlinear adaptive filtering.

We will look at the different goals and see how they are accomplished.

Before treating adaptation mechanisms the design of linear optimal filters is discussed. This gives the reader a good background and makes the book self-contained. When using the book as a text book this can be a disadvantage, since the basic material may be covered in a different course based on a different text book. On the other hand it gives the background and the notations that are needed in the adaptive part of the book.

The first goal is the main part of the book (even if it does not start until p. 273). The linear FIR adaptive filters are divided into: gradient algorithms, recursive least-squares (RLS) methods, and fast RLS algorithms.

The gradient algorithms cover the well-known least mean square (LMS) algorithm, its advantages and disadvantages are discussed and illustrated. The computational burden is low when using LMS and other algorithms based on stochastic gradients. The drawback is the slow rate of convergence. The next logical step is then to trade convergence rate against increased amount of computations. Different ways to organize the computations for RLS algorithms are discussed, e.g. QR-decomposition and Givens rotation. The main part of the computations in the RLS algorithms is the updating of a matrix. This leads to a computational complexity that increases with M^2 where M is the order of the filter. The LMS algorithm increases with M. The last algorithms have the property that the computational complexity also increases with M. The fast algorithms are more complex than the LMS algorithms, but have the same convergence properties as the RLS algorithms.

The material on design of linear FIR adaptive filters covers about 400 pages. The treatment is logical and thorough and leads the reader carefully towards more and more complex algorithms. The advantages and disadvantages of the different algorithms are discussed in great detail. The author also gives many hints and interpretations of the different algorithms. In my opinion the first goal of the aim is very well achieved.

The properties of the different algorithms are illustrated by computer simulations. All the details of the simulations are not always given. This makes it difficult for the reader to recreate the simulations and test the influence of different parameter choices. The examples are, however, well chosen and give good insight into the properties of the different algorithms. In my opinion there could have been more simulations of the different aspects of the algorithms. This could have been done at the expense of the description of some of the algorithms.

Nonlinear filtering and especially nonlinear adaptive filtering is a difficult area. This is treated in one of the last chapters of the book. The problem of blind deconvolutions is discussed. The problem is to reconstruct an unknown signal after it has been filtered through an unknown system. The problem of blind deconvolution is important and have many

* *Adaptive Filter Theory*, 2nd ed. by Simon Haykin. Prentice Hall, Englewood Cliffs, NJ (1991). ISBN 0-13-013236-5.

practical implications. To solve the problem it is necessary to assume that the desired signal is non-Gaussian and use the higher order statistics of the measured signal. This leads to nonlinear estimation. This part of the book is far outside the scope of the rest of the book and could have been omitted.

Numerical properties such as finite precision effects and numerical stability are discussed and illustrated in the end of the book. These are important problems that often are neglected in text books.

Each chapter ends with a selection of problems. Working through the problems gives insight into the properties of the different algorithms. The problems often lead to new versions of the algorithms and help the student to get a better understanding of the underlying theory. According to the preface there is a solutions manual available through the publisher. The solutions manual presents detailed solutions to all the problems in the book. Most chapters also have computer-oriented problems. These problems involve simulations of different algorithms and investigation of different choices of parameters. The use of computer simulations is important and will encourage the students to use their creativity and will increase the understanding of the methods.

The book has an extensive bibliography and the references are up-to-date. The author also has good descriptions of the historical development in the area of adaptive filtering.

In summary the book fulfills its goals with respect to the author's three aims. The book can be used as a text book in a graduate course in adaptive signal processing. It can also be used as a reference book for anyone who wants to find details about a number of algorithms for linear FIR adaptive filters. The book is strongly recommended for its wide coverage and for its depth.

About the reviewer

Björn Wittenmark was born in 1943. He received his M.Sc. degree in Electrical Engineering in 1966 and Ph.D. degree in automatic control in 1973, both from Lund Institute of Technology. In 1970 he became senior lecturer and in 1989 Professor in automatic control at Lund Institute of Technology. He is a Fellow of IEEE and has written numerous papers in the areas of adaptive control and digital control. He has co-authored the books *Computer Controlled Systems* and *Adaptive Control*.

Computational Methods for Linear Control Systems*

P. Hr. Petkov, N. D. Christov and M. M. Konstantinov

Reviewer: ANDRAS VARGA
Ruhr University Bochum, Department of Mechanical Engineering, Institute for Automatic Control, P.O Box 102148, D-4630 Bochum 1, Germany.

THE AIM OF THIS book is "to give a systematic presentation of methods oriented towards the computer aided analysis and design of linear control systems". Being the first book devoted exclusively to this topic, it fills an important gap in the existing control literature. The book deserves the important role of being a well documented guided introduction in the numerical aspects of solving system analysis and synthesis problems, and therefore it will be certainly useful for many control specialists unfamiliar with numerical techniques. Simultaneously, it represents a valuable reference for further research in the area of developing numerical algorithms for control systems.

The book illustrates the close interaction in the last two decades between two fields of applied mathematics: system theory and numerical analysis, and the emergence of a new, interdisciplinary field, which we can denote as the computational system theory. Specific demands for solving system analysis and design problems stimulated many researches in the area of numerical linear algebra. The algorithms developed for solving linear matrix equations (e.g. Lyapunov or Sylvester), for computing matrix functions (for instance the matrix exponential), or for ordering standard or generalized Schur forms by means of orthogonal transformations, all have strong appeals to important control problems. On the other side, the development of computational methods for linear systems was strongly influenced by the progress in the field of numerical algebra and the availability of high-quality software for solving linear algebra problems. Many of the sophisticated algorithms of linear algebra served as tools or as models for developing reliable algorithms for control problems. A notable example is the wide usage of the reduction of a square matrix to the real Schur form as a preliminary preprocessing step in many proposed algorithms. Virtually for almost all main computational problems of linear systems theory, algorithms labeled as "Schur" methods are available or can be derived (although they are not always the best of possible approaches).

The contents of the book may be outlined as follows. Chapter 1 deserves the role of introducing the basic techniques of numerical matrix computations. Several introductory paragraphs are devoted to basic numerical issues: rounding errors, conditioning of problems, numerical stability, performances of algorithms. The rest of the chapter is devoted to the main topics of numerical linear algebra: the solution of linear equations and the computation of eigenvalues. Along with presenting solution techniques for these and related problems, algorithms for computing various matrix decomposition (LU, Cholesky, QR, Hessenberg, Schur and singular value decomposition (SVD)) are described. This chapter clearly demonstrates the necessary emphasis on using orthogonal transformations in developing numerically reliable algorithms. Orthogonal transformations produce small, easy to bound errors, when they are applied to other matrices. Algorithms based exclusively on the use of such transformations have sometimes the highly desirable property of numerical stability, and therefore represent reliable numerical approaches for solving the respective computational problems. A more detailed presentation of the material presented in this chapter can be found in the book of Golub and Van Loan (1983).

Chapter 2 introduces the main concepts and reviews several fundamental results (without proofs) of the state-space analysis and design of continuous and discrete-time linear control systems. The computational problems to be addressed in the next chapters are formulated here and, for some problems, simple solution methods are given. The presented methods are, however, of purely theoretical

* *Computational Methods for Linear Control Systems* by P. Hr. Petkov, N. D. Christov and M. M. Konstantinov. Prentice Hall International (U.K.) Ltd. (1991). ISBN 0-13-161803-2.

interest and are not suitable for computer implementations. This aspect is illustrated several times later in the book.

Chapter 3 is devoted to the solution of state equations by using methods based on matrix exponentials. After presenting the main results on the sensitivity of matrix exponential, three classes of numerical approaches for its evaluation are presented: power series methods, ordinary differential equation methods and methods based on matrix decompositions. Well chosen examples illustrate the merits and drawbacks of different approaches. The accompanying error analyses provide useful bounds for assessing the precision of results. The best general purpose method to compute the matrix exponential is (probably) a combination between the block-diagonalization technique (Algorithm 3.4) and the Padé approximation method (Algorithm 3.2). A word of caution concerning the presented series methods (Algorithms 3.1 and 3.2) is necessary here. Both methods use a balancing procedure (described vaguely in Section 1.10) intended to reduce the 1-norm of the given matrix **A**. Although no reference is cited in the book, usually the popular balancing method implemented in the Fortran subroutine BALANC form EISPACK is assumed to be used. However by using this routine, the 1-norm of the given matrix may occasionally increase (sometimes drastically) preventing thus the applicability of methods due to unavoidable overflows. This aspect is apparently overlooked in several existing implementations of the Padé method and spectacular failures can occur even for simple examples as the following one

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & \epsilon & 1 \end{bmatrix}$$

where ϵ is a very small quantity (for example, the machine relative precision).

Chapter 4 discusses algorithms for solving computational problems related to the analysis of linear control systems. Several methods are described for the analysis of stability of linear systems and for studying the robustness of this property in presence of parametric perturbations. Then, the closely related problems of solving Lyapunov and Sylvester matrix equations are considered. The perturbation analysis of the Sylvester equation precedes the presentation of the Bartels–Stewart algorithm for its solution. The analysis of roundoff errors of this algorithm reveals its numerical stability. Variants of this algorithm for solving continuous and discrete Lyapunov equations or to compute directly the Cholesky factor of the non-negative definite solution of a class of Lyapunov equations are also presented. A more efficient variant of the Bartels–Stewart algorithm, known as the Hessenberg–Schur method, is also presented. The methods for analysis of controllability and observability properties are based on a numerically stable algorithm for computing the so-called controllability (or observability) staircase canonical form of a system. A proof of the numerical stability of this algorithm is given. The chapter ends with a discussion of methods for system balancing and model reduction.

Chapters 5 and 6 concern with the design of control systems by state-space methods. Chapter 5 presents computational methods for pole assignment of linear systems. Two numerically stable Schur methods are described. Both methods are based exclusively on the use of orthogonal transformations and produce, besides the state feedback matrix assigning the desired poles of the closed loop system, the Schur form of the closed-loop systems state matrix and the corresponding orthogonal transformation matrix. Both methods are suitable to design observers by pole assignment. The observer state matrix can be directly determined in a Schur form, a useful feature for increasing the efficiency of on-line computations when the observer operates in real-time. Robust pole assignment techniques are also considered. These methods try to assign simultaneously with the poles also a well conditioned set of eigenvectors of the closed-loop systems state matrix. For discrete systems, a special purpose pole assignment algorithm is presented for designing minimum norm dead-beat controllers. Finally, several methods are described which place directly the closed-loop poles in the stable region of the complex plane, without however involving an explicit or exact assignment of these poles. In spite of their desire to present the best of existing methods, the authors apparently overlooked the stabilization algorithms proposed in Varga (1981) which, in my opinion, are more suitable to solve reliably the stabilization problem than the methods presented by the authors.

Chapter 6 is devoted to computational methods for solving continuous and discrete-time algebraic matrix Riccati equations. Both iterative (Newton's and matrix sign function) methods as well as direct (Schur vectors, generalized Schur vectors and symplectic transformation) methods are presented. Interesting examples are given which illustrate various aspects, as for example: the sensitivity of problems, numerical stability issues, or the accuracy of methods. Worth mentioning are Examples 6.7 and 6.9 showing that generally the Schur vectors methods cannot be unconditionally considered as numerically stable methods. In comparing the relative efficiencies of Newton's and Schur methods, the authors arrive to a somewhat misleading conclusion that Schur methods are generally *much* more efficient than Newton's methods. This conclusion is based on a overly pessimistic operations count for the Newton's method ($200n^3$ *flops* for 10 iterations) and too optimistic operations counts for the Schur method ($75n^3$ *flops*) and generalized Schur method ($180n^3$ *flops*). Actually, by using the authors' evaluations to compute the Schur form (about $13n^3$ *flops*) and the generalized Schur form (about $33n^3$ *flops*), we obtain the following figures: about $150n^3$ *flops* for ten Newton iterations, about $130n^3$ *flops* for the Schur vectors method and about $320n^3$ *flops* for the generalized Schur vectors method. However, the Newton's method usually converges to the limiting accuracy solution in at most seven–eight iterations, and therefore this method has usually the same efficiency as the Schur vectors methods. In passing we also note that all memory requirement evaluations for various algorithms are in excess with at least $2n^2$ storage locations. The implementations listed in Varga and Sima (1993) illustrate the possibility to implement these methods with lower memory usage.

Chapter 7 presents methods for computing with subspaces (sum, intersection, image, pre-image and angles) and for determining various subspaces (controllable, unobservable, controlled invariant and controllability subspaces) arising in the geometric theory of linear systems. The algorithms which perform such computations are based on reliable numerical techniques to assess the rank of matrices (SVD or QR decomposition with column pivoting). An interesting application of the presented techniques is the computation of the Kalman decomposition of a linear system. Many examples are used to illustrate the main computational aspects.

We can formulate several closing remarks. The book offers a comprehensive overview of existing numerically reliable algorithms for solving the basic systems analysis and design problems. The selection of presented algorithms reflects clearly the important progress made in the last two decades in developing numerically reliable algorithms for control systems and in studying the sensitivity of various systems analysis and design problems. The authors themselves have several notable contributions to this progress. The performed error analyses (also partly contributions of authors to the field) are useful premises for further research. Carefully chosen examples help to clarify many of discussed aspects and contribute decisively to the understanding of merits or inconveniences of particulars methods. The exercises given at the end of each paragraph highlight additional aspects not discussed in the book. Some topics of possible interest omitted from the presentation are: the computation of systems zeros, the computation of frequency response, or the evaluation of transfer function matrices from state-space models.

The algorithms presented in the book can serve in principle for computer implementations, although this

aspect, in my opinion, was not the main intention of authors in presenting the algorithms. The methods for analysis and design of linear systems are rather sophisticated and therefore, the development of robust software implementations is a task for numerical analysis experts. Many algorithmic details, not mentioned in the book can further improve the efficiency of algorithms and the accuracy of results. For almost all of presented methods and for many other algorithms only referred to, robust implementations are available either in Fortran libraries or in MATLAB Toolboxes.

References

Golub, G. H. and C. F. Van Loan (1983). *Matrix Computations*. The John Hopkins Univ. Press, Baltimore.

Varga, A. (1981). On stabilization algorithms for linear time-invariant systems. *Rev. Roum. Sci. Techn.—Electrotechn. & Energ.*, **26**, 115–124.

Varga, A. and V. Sima (1993). *Algorithms and Commented Fortran Subprograms for Computer Aided Control Systems Engineering*. Technical Press, Bucharest. (In print).

About the reviewer

Andras Varga was born in 1950 in Baraolt, Romania. He obtained the diploma in control engineering in 1974 and the Ph.D. degree in electrical engineering in 1981, both from the Polytechnic Institute of Bucharest. In 1974, Dr Varga joined the Institute of Informatics in Bucharest where he worked first as a research and development engineer, and beginning with 1982 as a senior research engineer. In 1990 he was promoted as first degree senior research engineer, a research position equivalent with the academic position of full professor. In 1987, Dr Varga was awarded with an Alexander von Humbold Research Fellowship. Since September 1990, he has been working at the Department of Mechanical Engineering of the Ruhr University of Bochum in the framework of this fellowship.

Dr Varga's interests are in numerical methods for control systems, with emphasis on model reduction, robust control and descriptor systems, as well as in the area of developing software tools (libraries, interactive packages) for the computer aided design of control systems (CADCS). He has authored or coauthored more than 50 papers and three books (two of them being in print). He is the author or coauthor of several professional level CADCS packages and of a H_∞-synthesis MATLAB Toolbox. Dr Varga is member of the Benelux Working Group on Software and an active contributor to the SLICOT Control Library developed by this group.

Decentralized Control of Complex Systems*

D. D. Siljak

Reviewer: RENHOU LI
Xian Jiatong University, Xian, P.R. China.

MANY REAL ENGINEERING and non-engineering problems facing researchers of the world are highly complex and stochastic in nature. Modelling and controlling of most complex systems must deal with 'high' dimension, uncertainty, and information structure constraints. One way for solving these problems is to use the strategy called decentralized control in which system inputs are assigned to a given set of local controllers (stations) which observe only local system outputs. This approach can avoid difficulties in data gathering, storage requirements, computer program debugging and geographical separation of system components. A great number of papers and books (Jamshidi (1983); Siljak (1978); Singh (1981), etc.) have been published to indicate concepts and methodologies of decentralized control of complex system since the 1970s. Decentralized control has been successfully applied to many industrial systems, for example, power systems, traffic systems and communication systems, and has become one of the most important branches of modern systems theory.

Following the book *Large Scale Dynamic Systems: Stability and Structure* which presented some ideas about decentralized control, published in 1978, Professor Siljak has written another book *Decentralized Control of Complex Systems* which is mainly concerned with the characteristics of complex systems: dimensionality, uncertainty, and information structure constraints, and provides a series of novel concepts and methods to handle the problems of decentralized control of interconnected subsystems.

The book summarizes the research achievements developed by the author and his colleagues in the decentralized control field in the last two decades. Throughout the book great care is taken to differentiate between centralized and decentralized control. In this book, a great deal of attention is paid to the explanation of the substantial problems involved in decentralized control systems, such as interconnected subsystems modelling, decomposition, stability, optimization and robustness of decentralized controllers. In the first chapter of the book the important concept of graph-theoretic framework is presented starting from the view point of structural modelling for complex systems. Because the decentralized control problems are essentially structural, the graph-theoretic framework provides a very good environment for the computation of system structural fixed modes with arbitrary feedback structure constraints, reachability, controllability and observability. It also can be used to identify the minimal set of lines in directed graphs which are essential for preserving input reachability and structural controllability of a given system. The graph-theoretic algorithm is computationally attractive because one can use the Boolean operations instead of algebraic manipulation to get results needed. I think that the graph-theoretic framework and its algorithms which are used repeatedly, almost in each chapter, can be considered as the cream of the book, it is an excellent part, which is superior to other books on the same topic.

In the second chapter of the book, an effort is made to deal with the important problem of the connective stability of complex systems under structural perturbations. The book provides an M-matrix condition based on Vector Lyapunov Function for testing the connective stability of decentralized control systems. Having this result, the control strategy may become very simple: stabilize each subsystem when decoupled, and then check stability of the connective closed-loop subsystems using M-matrix conditions. The condition is further modified in the successive chapters and frequently applied to study inherent properties of decentralized control systems, such as robustness, suboptimization and decentralized stabilizability of interconnected subsystems. The entire second chapter should be considered as the theoretical fundamentals of the later chapters.

The third part of the book is concerned with the optimization of decentralized control systems. Due to the

* *Decentralized Control of Complex Systems, Math. en Science and Eng.*, Vol. 184 by D. D. Siljak. Academic Press, New York (1991). $75.00.

nonclassical information structure constraints, the standard optimization concepts and methods cannot be extended to formulate decentralized control strategies. The most important factor to be considered is the interconnection between the subsystems. In this context, the book regards interconnections as perturbations of subsystems that are controlled by decentralized LQ feedback and introduces a new idea of suboptimality index to measure the cost of robustness to structural perturbations. The book constructs and solves the inverse optimal problem of decentralized control which leads to an interesting result that the global optimality can be recovered from local optimal LQ control subsystems (even with nonlinear interconnections) if the performance index is modified. This is a very useful conclusion for the design of decentralized control laws, and also is an important contribution to the theory and application of decentralized control.

Based on the graph-theoretic framework, interpreting the interconnections as perturbations of subsystems and using **M**-matrix test techniques the book extend the main results made in Chapters 1–3 to Chapters 4 and 5, in which stochastic control, dynamic feedback control, and adaptive control of complex systems are analysed and discussed. In these parts an effort is made to derive many important theorems which establish the basis of arguments, such as the sufficient conditions for the existence of decentralized asymptotic observers, suboptimal estimator, and optimal decentralized control laws for a given system. The book provides several methods for designing decentralized observers, estimators and controllers which can be carried out by using a parallel processing scheme.

In order to control a complex system decentrally, the prerequisite is that the system can and must be decomposed into several subsystems. But how is this done? The book gives sufficient space (three chapters) to introduce three decomposition approaches: LBT (Low Block Triangular) decomposition, nested Epsilon decomposition and overlapping decomposition, which can be easily achieved using the algorithms listed in the Appendix demonstrating the inherent efficiency of these algorithms and their easy implementation in a structural language, such as PASCAL and C.

The last part of the book is concerned with the reliability of control which is the basic requirement in design of complex systems. So far as reliability is concerned, it has been discussed from different view points in relevant disciplines and there exist various techniques to increase the reliability of control systems from both hardware and software. In the book, the emphasis is placed on the control structure and multiple control systems. Using the Inclusion Principle and overlapping decomposition made in the preceding chapter it gives an example to illustrate the design of reliable control. The contents discussed in the last chapter are not detailed and may be regarded as a brief introduction to the reliable control of complex systems.

Finally it is worthy to point out that at end of each chapter the book includes a section of notes and references, which not only provide readers with a large amount of relevant literature, but also introduce the evolution of some novel concepts and methodology obviously beneficial to those who wish further to pursue studies in some subjects of decentralized control of complex systems. Another merit of the book is that each chapter contains many examples of both analysis and design which makes the content of the book to be understood easily and well-readable. In my opinion the book is tailored to the needs of the readers who have acquired a degree of proficiency in modern control systems theory.

References

Jamshidi, M. (1983). *Large-scale Systems: Modelling and Control.* North-Holland, New York.

Siljak, D. D. (1978). *Large-scale Dynamic Systems: Stability and Structure.* North-Holland, New York.

Singh, M. G. (1981). *Decentralized Control.* North-Holland, New York

About the reviewer

Renhou Li graduated from Jiaotong University, Shanghai, China, in 1957, he began working in Xian Jiaotong University as a teaching assistant, lecturer, associate professor and now as a professor. He was the vice chairman of the Department of Radio Engineering from 1978 to 1984. Since then he has been the Dean of ICE department. His main research interests are in Large Scale Systems Theory and Its Applications, Distributed Computer Control Systems, The Application of Expert Methodology to Industrial Process and Intelligent Control. He has more than 60 papers published and is the author and coauthor of six books.

Automatica, Vol 29, No 2, pp 573–583, 1993
Printed in Great Britain
Pergamon Press Ltd

Biographical Notes on Contributors to this Issue

Tamer Başar was born in 1946, in Istanbul, Turkey. He received B.S.E.E. from Robert College, Istanbul, in 1969, and M.S., M.Phil, and Ph.D. degrees in engineering and applied science from Yale University, in 1970, 1971 and 1972, respectively. After being at Harvard University, Marmara Research Institute (Gebze, Turkey), and Bogaziçi University (Istanbul), he joined the University of Illinois at Urbana-Champaign in 1981, where he is currently a Professor of Electrical and Computer Engineering. He has spent two sabbatical years (1978–1979 and 1987–1988) at Twente University of Technology (The Netherlands) and INRIA (France), respectively.

He has authored or co-authored over one hundred journal articles and book chapters, and numerous conference publications, in the general areas of optimal control, dynamic games, stochastic control, estimation theory, stochastic processes, information theory, and mathematical economics. He is the co-author of the text *Dynamic Noncooperative Game Theory* (Academic Press, 1982), editor of the volume *Dynamic Games and Applications in Economics* (Springer-Verlag, 1986), co-editor of *Differential Games and Applications* (Springer-Verlag, 1988), and co-author of the text *H^∞-Optimal Control and Related Minimax Design Problems* (Birkhäuser, 1991).

Dr Başar carries memberships in several scientific organizations, such as Sigma XI, SIAM, SEDC, ISDG, and IEEE. He was elected a Fellow of IEEE in 1983, and has served its Control Systems Society in various capacities, most recently as an associate editor at large and the Editor for TNC for its technical journal, and as the general chairman of its major conference in 1992. He has also been active in IFAC, in the organization of several workshops and symposia, and more recently as an editor of *Automatica*. Currently, he is also the President of the International Society of Dynamic Games (ISDG), the Managing Editor of the Annals of ISDG, and an associate editor of the *Journal of Economic Dynamics and Control*.

J. Bernussou is Director de Recherche at CNRS-France. He is graduated from the ENSEEIH, Toulouse engineering school and took his Doctorate d'Etat es Sciences Physiques at University Paul Sabatier in 1974. He heads a research group in LAAS-CNRS, Toulouse. His field of interest concerns analysis and control of large scale systems, communication networks, distributed control and robust control of uncertain linear systems.

Sergio Bittanti received the *Laurea* in Electronic Engineering from the Politecnico di Milano and the *Laurea* in Mathematics from the Universita' degli Studi di Milano. Since 1971, he has been with the Dipartimento di Elettronica e Informazione of the Politecnico di Milano (Italy), where he teaches automatic control, identification and optimization and system theory, and coordinates the Ph D. program in 'computer and control engineering'.

He is one of the founding members of the Societa' Italiana di Matematica Applicata e Industriale, which was established in 1989 as the Italian branch of SIAM, a member of IFAC Committee on Theory and a member of ECCA (European Community Control Association). For five years, he has served as national coordinator of the research project model identification, system control, signal processing, a project funded by the Italian Ministry of University and Scientific Research which groups more than 100 researchers and professors of 15 Italian Universities.

Dr Bittanti is the author of many scientific papers and author or editor of a number of books, the latest of which is *The Riccati Equation* (1991, Springer-Verlag, Berlin with A. J Laub and J. C. Willems co-editors). He is an associate editor of *Computer and Statistics* and the *International Journal of Mathematical Systems, Estimation and Control*.

His current scientific interests focus on PSICO (Periodic Systems Identification, Control and Optimization) and adaptive identification for adaptive control. He is involved in various application projects, such as active control of vibration and noise in transportation systems, neurocontrol and computer vision

---

Yossi Chait was born in Ramat-Gan, Israel in 1957. He received a P.M.E. from Ort Technical College, Israel in 1977, a B.S. in mechanical engineering in 1982 from The Ohio State University, and an M.S. and Ph.D. in mechanical engineering from Michigan State University in 1984 and 1988, respectively. In 1985 he held a summer position at General Motors Research Labs. Since 1988 he has been an Assistant Professor at the mechanical engineering department at the University of Massachusetts, Amherst, MA. He is a member of ASME. He has authored and co-authored numerous papers on control of uncertain distributed parameter systems and on the quantitative feedback theory. His current research interests are theory and implementation of (robust) feedback systems.

Joe H. Chow received the B.S. degrees in electrical engineering and mathematics from the University of Minnesota, MN, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana, for his work on singular perturbations in control system design. From 1978 to 1987 he worked at the General Electric Company on power system dynamics and control problems. He is currently Professor of Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute, Troy, New York. His interests include multivariable control systems, large-scale systems, and power systems. Dr Chow is an Associated Editor of *Automatica*. He is a fellow of IEEE, and is currently a member of the Board of Governors of the Control Systems Society. He was also a recipient of the Donald P. Eckman Award.

Giuseppe Conte was born in Genova, Italy, on 30 September 1951. He received the Doctor degree in Mathematics in 1974 from the University of Genova. He has held the position of lecturer, from 1974 to 1985 and Associate Professor, from 1985 to 1990, at the University of Genova and is now Professor of System Theory at the University of Ancona, Italy. He was Fulbright Scholar in 1980 and 1987 and holds a NATO Senior Fellowship in 1987. He has been Visiting Professor at the Ohio State University of Columbus, OH, in 1987 and at the Ecole Normale Superieure de Mechanique de Nantes, France, in 1988. His research interests concern algebraic and geometric methods in system and control theory Dr Conte is an Associate Editor of the *SIAM Journal on Control and Optimization*. He is member of IEEE, A.M.S., and U M I.

Martin Corless is currently an Associate Professor in the School of Aeronautics and Astronautics at Purdue University, West Lafayette. He received a B.E. from University College, Dublin, Ireland and a Ph.D. from the University of California at Berkeley; both degrees are in mechanical engineering. He is the recipient of a 1990 National Science Foundation Presidential Young Investigator Award.

His research is concerned with obtaining tools which are useful in the robust analysis and control of systems containing significant uncertainty and in applying these results to aerospace and mechanical systems.

Mark J. Damborg received the B.S. degree in electrical engineering from Iowa State University and the M.S. and Ph.D. degrees from the University of Michigan in 1963 and 1969, respectively. He spent the academic year 1966–1967 as a Fulbright Fellow at the Technological University in Delft, The Netherlands. Since 1969 Dr Damborg has been at the University of Washington where he is now Professor of Electrical Engineering. His research interests concern analysis and control of dynamical systems with emphasis on adaptive and optimal control concepts and control of power systems. Associated interest include database applications to computer-aided engineering systems.

Ranjit A. Date was born in Pune, India He was awarded the Gold Medal when receiving his Bachelor's degree in Instrumentation and Control Engineering from the University of Pune in 1987 He received his M.S. degree in Computer and Systems Engineering from Rensselaer Polytechnic Institute, Troy, NY in 1989. He received a Ph D., also from RPI, for his work in Decentralized Control Theory in 1991 He has worked with System Dynamics group at the General Electric Company for the development of dynamic equivalencing programs for large power systems He is a founder and presently the Executive Director of Precision Automatica and Robotics India P. L , Pune, India His areas of interest include Decentralized Control Theory, implementational issues in control systems, and Computer-aided Control System Analysis and Synthesis.

Edward J. Davison was born in Toronto, Canada in 1938. He received the A.R C.T. degree in piano from the Royal Conservatory of Music of Toronto in 1958, the B.A.Sc. degree in Engineering Physics and the M.A. degree in Applied Mathematics from the University of Toronto in 1960, 1961, respectively. In 1964 he received the Ph.D. degree, and in 1977 the Sc.D. degree from Cambridge University, U.K.

From 1964 to 1966 he was with the University of Toronto; in 1966–1967 he was with the University of California, Berkeley, in the Department of Electrical Engineering and Computer Science, and since then, he has been with the Department of Electrical Engineering, University of Toronto. His current research interests include the study of multivariable control system theory and design, large scale systems, robust control, and adaptive control.

Dr Davison was Associate Editor from 1974 to 1976, Guest Associate Editor in 1977–1978, 1982–1983, and Consulting Editor in 1985, of the *IEEE Trans. on Automatic Control*. He was an Associate Editor of *Automatica* from 1974 to 1987, and of *Large Scale Systems: Theory and Applications* from 1979–1990. He was Vice-President (Technical Affairs) in 1979–1981, President-Elect in 1982, and President in 1983, of the IEEE Control Systems Society. He was Vice-Chairman of the International Federation of

Automatic Control (IFAC) Theory Committee in 1978–1987, Chairman of the IFAC Theory Committee in 1988–1990, and at present is serving as Vice-Chairman of the IFAC Technical Board and a member of the IFAC Council since 1991. He received an Athlone Fellowship in 1961–1963, the E.W.R. Steacie Memorial Fellowship in 1974–1977 and the Killam Research Fellowship in 1979–1980; 1981–1983. In 1984, he received the IEEE Centennial Medal and was elected a Distinguished Member of the IEEE Control Systems Society. He was elected a Fellow of the Royal Society of Canada in 1977, and a Fellow of the IEEE in 1978; in 1986, he was elected "Honorary Professor" of Beijing Institute of Aeronautics and Astronautics. He is a designated consulting engineer of the Associate of Professional Engineers of the Province of Ontario since 1979, and a Director of Electrical Engineering Consociates Ltd. He has received a number of Best Paper Awards from the *IEEE Trans. on Automatic Control*, and has a Current Contents Classic Paper Citation paper. He is currently serving on the editorial board of a number of journals.

Giuseppe De Nicolao was born on 24 April 1962, in Padova, Italy. He received the *Laurea* in Electronic Engineering in 1986 from the Politecnico of Milano, Italy. From 1987 to 1988 he was with the Biomathematics and Biostatistics Unit of the Institute of Pharmacological Researches "Mario Negri", Milano. In 1988 he entered the Italian National Research Council (C.N.R.) as a researcher of the Centro di Teoria dei Sistemi in Milano. In November 1992 he joined the Dipartimento di Informatica e Sistemistica of the Università di Pavia (Italy), where he teaches model identification. In the summer of 1991, he has held a visiting fellowship at the Department of Systems Engineering of the Australian National University, Canberra (Australia).

His research interests include linear periodic systems and cyclostationary processes, multirate sampled-data control systems, Riccati equations, deconvolution techniques, modeling and simulation of endocrine systems

Guang Ren Duan was born in Heilongjiang Province, P. R. China on 5 April 1962. He received the B.S. degree in Applied Mathematics from the north-east Heavy Machinery Institute, Qiqihar, P. R. China, in 1983 and the M. S. degree in Modern Control Theory from Heilongjiang University, Harbin, P. R. China, in 1986. Subsequently, he received the Ph.D. degree in General Mechanics from Harbin Institute of Technology, Harbin, in 1989. And afterwards, he worked as a post Ph.D. graduate at Harbin Institute of Technology for two years.

He is presently a Professor in the department of Control Engineering at Harbin Institute of Technology and a member of IEEE Control Society. His current research interests are in the areas of linear system theory, singular systems and robust control.

Procopis Fessas was born in Thessaloniki, Greece, in 1950. In 1974 he obtained the Diploma in electrical engineering from the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland, and in 1979 the Doctor's Degree (Dr. Sc. Techn.). In 1980 he joined the Aristotle University of Thessaloniki, where he currently has the position of an Associate Professor. He has been several times at the Institute of Automatic Control of ETH-Zurich as an academic guest and as a visiting Professor. In 1989 he was at the Laboratoire d'Automatique (LAAS) of the CNRS in Toulouse, France, during his sabbatical year. His research interests are control of large scale systems, decentralized control of interconnected systems, control and transient stability of electrical power systems. He is the author of several papers in the above areas.

James S. Freudenberg was born in Gibson County, IN, on 31 March 1956. He received the B.S. degree in mathematics and physics from the Rose-Hulman Institute of Technology, Terre Haute, IN, in 1978, and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, in 1982 and 1984, respectively.

From 1978 to 1979 he worked for the Navigation Systems Technology Group, Rockwell-Collins Avionics, Cedar Rapids, IA. He held a Research Assistantship with the Coordinated Science Laboratory at the University of Illinois, Urbana-Champaign. In 1982 he held a summer position with the Systems and Research Center, Honeywell, Minneapolis, MN. He is currently an Associate Professor in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI.

Francesco Garofalo was born in Naples on 14 January 1952. He received the "Laurea" degree in electronics engineering in 1976 from the University of Naples. In 1979 he was appointed a Research Associate at the System Theory Center, Milan Polytechnic; he moved to the University of Cosenza in 1980 and to the Naval Academic Institute of Naples in 1982. In 1985 he joined the department of Computer and Systems Engineering, University of Naples, as an Associate Professor of Automatic Control. He was a visiting scholar at the School of Mechanical Engineering, University of California, Berkeley in 1987.

His interests include the analysis and control of systems via Lyapunov methods, control of singularly perturbed systems, control of plasma position in tokamak reactors, and diving.

José C. Geromel was born in Itatiba, Brazil, in July 1952. He received the B.Sc. and M.Sc. Degrees in Electrical Engineering from the State University of Campinas (UNICAMP), Campinas, Brazil, in 1975 and 1976, respectively, and the Docteur d'Etat es Sciences Physiques degree from the University Paul Sabatier, Toulouse, France, in 1979. In 1975 he joined the Faculty of Electrical Engineering of the State University of Campinas where he is presently Professor of Control Theory and Systems Analysis. In 1987 he was a visiting professor in the Departimento di Elettronica of the Instituto Politecnico di Milano, Milan, Italy.

Luigi Glielmo was born in Benevento, Italy, on 7 April 1960. He received the "Laurea" degree in electronics engineering and the Research Doctorate in automatic control from the University of Naples, Italy. In 1990 he joined the Department of Computer and Systems Engineering, University of Naples, as a researcher. In 1992 he was promoted to Associate Professor of Automatic Control. In 1989 and 1990 he was a visiting scholar at the School of Aeronautics and Astronautics, Purdue University, Indiana and, in 1990, a visiting scientist at the NET Team, Max-Planck Institut fur Plasmaphysik, Garching bei Munchen, Germany.

His interests include singular perturbation methods, analysis and control of systems via Lyapunov methods, robust stability, identification of plasma shape in tokamak reactors, and sailing. He is a member of IEEE and SIAM.

Alena Halouskova graduated as electrical engineer from the Czech Technical University, Prague, in 1957. In 1965 she received the Ph.D. degree in control engineering from the Czechoslovak Academy of Sciences. Since 1960 she has been with the Institute of Information Theory and Automation mostly working in the field of adaptive control. Design and applications of adaptive control in industrial systems are her current major research interests.

Kenneth Hunt is from Glasgow, U.K. He is currently supported by a Personal Research Fellowship of the Royal Society of Edinburgh and is a member of the Control Group in the Department of Mechanical Engineering at the University of Glasgow. He joined the Group in 1989. In November 1992 he joins the research institute of Daimler-Benz AG in Berlin.

Dr Hunt obtained a B.Sc. in Electrical Engineering in 1984 and a Ph.D. in Control Theory in 1987, both from the University of Strathclyde (Glasgow's second university). At Strathclyde, he was employed as a Research Assistant funded by Scottish Power (formerly SSEB) and developed optimal and self-tuning control algorithms for power systems applications. From 1987 to 1989 he was a scientist with BBN Systems and Technologies (the European division of Bolt, Beranek and Newman Inc.) in Edinburgh, he worked on Expert Systems for process monitoring, intelligent decision support systems for manufacturing applications, and on parallel processing.

His current research areas include adaptive control, the polynomial equation approach to optimal control and the application of connectionist architectures to nonlinear control problems. Recent forays into AI have produced interesting results in control applications of machine induction. Currently, he is applying genetic algorithms to control system optimization.

Kenneth Hunt is the author of the monograph *Stochastic Optimal Control Theory with Application in Self-tuning Control* (1989, Springer-Verlag, Berlin). A Chartered Engineer, he is a member of the Institution of Electrical Engineers and serves on the Professional Group Committee on Control and Systems Theory.

Robert G. Hutchins was born in Asheville, North Carolina. He received an A.B. degree in History from Oberlin College and B.S. and M.S. degrees from the University of Florida in Chemistry and Statistics, respectively. In 1988 he received the Ph.D. degree in Systems Science from the University of California, San Diego. Between 1978 and 1982 he was employed as an Operations Research Analyst for the Department of the Navy. From 1982 until 1989 he was employed as a Systems Engineer at TRW in San Diego. In 1989 he joined the ORINCON Corporation in San Diego where he specializes in signal identification, data correlation and tracking.

Dr Hutchins is interested in the filtering and control of stochastic systems and in related issues such as signal and system identification and sensor fusion. At TRW he helped develop sensor fusion and target tracking algorithms for the TRW/Westinghouse integrated electronic warfare system (INEWS). His responsibilities at ORINCON involve the development of multisensor detection, recognition and tracking systems for ASW. He is also pursuing research interests in imaging systems and their role in multisensor trackers and in the use of neural nets for signal and system identification.

Altuğ Iftar was born in Istanbul, Turkey on 28 June 1960. He received the B.S. degree in electrical engineering from Boğaziçi Universitesi, Istanbul, Turkey in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the Ohio State University, Columbus, OH, U.S.A. in 1984 and 1988 respectively. He held teaching assistantship positions at Boğaziçi Universitesi from 1980 to 1982 and graduate teaching and research associateship positions at the Ohio State University from 1983 to 1988. He was with the Department of Electrical Engineering at the University of Toronto, Toronto, Canada

from 1988 to 1992. He joined the Department of Electrical and Electronics Engineering at Anadolu Üniversitesi, Eskişehir, Turkey as an Associate Professor in 1992. His research interests include large scale systems, decentralized control, robust control, optimization, and optimal control.

Alan Johnson was born in London in 1946. He graduated in 1967 from Birmingham University as a chemical engineer and obtained his Ph.D. in control engineering from Cambridge University in 1970. After working with Shell in the U.K. and The Netherlands he joined Delft University of Technology in 1975, and was appointed to his present position as Professor of Process Control and Dynamics in 1980.

Professor Johnson has consulted for various companies, and is a member of the board of scientific directors of a research institute. He is the author of a book on process control and has edited several proceedings of symposia, among which that of the first IFAC symposium on modelling and control of biotechnological processes. He is a member of the IFAC Working Group on Control and Instrumentation of Biotechnological Processes.

His current research interests are all aspects of LQG control, especially applications. In his spare time he plays cricket.

Miroslav Kárný studied theoretical cybernetics at the Faculty of Engineering Sciences and Nuclear Engineering, Czech Technical University Prague, and graduated in 1973. Since 1976, he has been with the Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, where he received the Ph.D. degree in 1978 and Dr. Sc in 1990. He has participated in developing theory and algorithms of Bayes-based approach to system identification and adaptive control as well as in variety of practical applications of these. Exploitation of prior information for pretuning of adaptive controllers and non-linear estimation are his main current interests.

Mark Kent received his B.A. in Applied Mathematics in 1985 and his M.S. in Systems Engineering in 1990 from the University of California, San Diego (UCSD). He is currently a Ph.D. candidate in Electrical Engineering at UCSD; his dissertation topic is 'Nonlinear Filtering and Image Processing'.

Mr Kent is currently a Principal Engineer at ORINCON Corporation, where he is responsible for developing an offboard target report processor to perform multiple hypothesis fusion on incoming data links for ORINCON's real-time display project for the Naval Underwater Systems Center (NUSC), Newport, RI.

Prior to joining ORINCON, Mr Kent was a member of the technical staff at Computer Sciences Center (CSC) in San Diego, CA. Here, he analysed the Navy's joint tactical information distribution system navigation performance in various operational environments.

Petar V. Kokotović received his graduate degrees from the University of Belgrade, Yugoslavia, in 1962, and from the Institute of Automation and Remote Control, Russian Academy of Sciences, Moscow, in 1965. From 1966 until March 1991, he was with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory at the University of Illinois, Urbana, where he held the endowed Grainger Chair. In April 1991 he joined the Electrical and Computer Engineering Department of the University of California, Santa Barbara, as a co-director of the newly formed multidisciplinary Center for Control Engineering and Computation. in 1990 he received the IFAC's Quazza Medal—the federation's highest award that has been given triennially since 1981. IFAC cited Kokotović for his outstanding research contributions to sensitivity theory, singular perturbations, and adaptive control. Petar Kokotović delivered the 1991 IEEE Bode Prize Lecture.

Konstantionos (Kostas) J. Kyriakopoulos was born in Athens, Greece on 19 September 1962. He received the Diploma in Mechanical Engineering with Honours from the National Technical University of Athens (NTUA), Greece in June 1985. In January 1986 he was enrolled in Rensselaer Polytechnic Institute where he received the M.S. Degree in December 1987 and Ph.D. Degree in May 1991, both in Computer and Systems Engineering.

During the years 1986–1988 he held various teaching and research assistantship positions. Between 1988 and 1990 he did research at the NASA Center of Intelligent Robotic Systems for Space Exploration. Currently, he is Research Assistant Professor in the Electrical, Computer and Systems Engineering Department at Rensselaer Polytechnic Institute associated with the New York State Center for Advanced Technology in Automation and Robotics, Rensselaer Polytechnic Institute. His research interests are in the area of nonlinear optimization and control theory with applications in robotic motion planning and control.

He is a member of IEEE and the Technical Chamber of Greece. He was awared the G. Samaras award of academic excellence from NTUA, the Bodosakis Foundation fellowship (1986–1989) and the Alexander Onasis Foundation fellowship (1989–1990).

Fotis N. Koumboulis was born in Athens, Greece, in 30 July 1964. He received the Diploma in Electrical Engineering and the Ph.D. degree in Automatic Control Systems from the National Technical University of Athens, in 1987 and 1991, respectively. Since 1988 he has been research associate at the NRC "Democritos", Institute of Information and Telecommunications, Athens, Greece. Since 1992 he has been special scientist at the NDRC, Department of Electronics, Athens, Greece. His research interests are in multivariable feedback control, mathematical system theory, adaptive control, optimal design techniques, aerospace control, and robotics.

Huibert Kwakernaak was born in 1937 in Rijswijk (Z.H.), The Netherlands. He obtained the diploma in Engineering Physics at Delft University of Technology in 1960. Following this, he studied Electrical Engineering at the University of California at Berkeley, where he obtained the M.Sc. degree in 1962 and the Ph.D. degree in 1963.

After returning to The Netherlands, Professor Kwakernaak first was a research associate and later a reader in the Departments of Engineering Physics and Mathematics of Delft University of Technology. Since 1970 he has been a full professor in the Department of Applied Mathematics of what is now the University of Twente, where he teaches systems and control theory.

Professor Kwakernaak's research interests are in linear control theory. He is the co-author of several books (*Linear Optimal Control Systems*, with R. Sivan, Wiley-Interscience, New York (1972); *Lineare Kontrolltheorie*, with H. W. Knoblock, Springer-Verlag, Berlin (1985); *Modern Signals and Systems*, with R. Sivan, Prentice Hall, Englewood Cliffs, NJ (1991)). Currently his research is focused on the "polynomial" approach to H_∞ optimization.

Huibert Kwakernaak is a Fellow of the IEEE and received the IFAC Outstanding Service Award.

J. B. Lasserre was born in France. He graduated from Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble (France) in 1976 and received the Docteur-Ingénieur et Docteur d'Etat degrees both from Université Paul Sabatier, Toulouse (France) in 1978 and 1984. During the academic years 1978–1979 and 1985–1986, he was a Postdoc INRIA and NSF, in the Electrical Engineering Department at the University of California at Berkeley. Since 1980, he has been with the Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS, Toulouse (France) He is currently Associate Editor of the *IEEE Transactions on Automatic Control* and the *International Journal on Production Research* His current research interests are in constrained linear systems, Markov decision processes and control theory applied to production planning and scheduling.

E. Bruce Lee received the B.S and M.S. degrees in mechanical engineering from the University of North Dakota, Grand Forks, in 1955 and 1956, respectively, and the Ph.D. degree from the University of Minnesota, Minneapolis in 1960.

From 1956 to 1963 he was employed by Honeywell, Inc. to develop digital inertial sensors and control system design techniques. Since 1963 he has been employed by the University of Minnesota, Minneapolis, where he is currently Professor of Electrical Engineering and Associate Director of the Center for Control Sciences and Dynamical Systems.

Yanlin Li received the B.S. degree in electrical engineering from the China Textile University, Shanghai, the People's Republic of China in 1982, and the M.S. degree in electrical engineering from the University of Minnesota, Minneapolis in 1989.

He is currently completing his Ph.D. degree in the Center for Control Science and Dynamical Systems at the University of Minnesota. His research interests include H^∞ control with applications, nonlinear control systems and traffic modelling and control.

Gengshen Liu was born in Tianjin, China in July 1945. He received the diploma in engineering from Harbin Engineering Institute, China in 1967. From 1967 until 1980 he did design and research work in the Chinese shipbuilding industry. He received M.S in naval architecture and marine engineering and Ocean Engineer degree in 1984, M.S in ocean system management in 1985, Sc.D. in ocean engineering in 1988, from the Engineering School of MIT and M.S. in management in 1990 from the Sloan School of Management of MIT

He had worked as research assistant, research associate at MIT in system identification and system simulation before he joined McDermott International Inc as a senior engineering analyst in September 1991 He is interested in mathematical modeling, system identification, system simulation, dynamic system and system dynamics.

He is the member of SIGMA XI

Jian-San Luo was born in TianJin, P R. China, in 1955 He obtained the B Sc degree in automatic control engineering from the Science and Technology University of National Defence, P. R. China in 1981. After working in industry for two and half years, he joined the Beijing Institute of Control Engineering of the Chinese Academy of Space Technology, and was awarded the M.Sc. in 1987 From 1987 to 1989 he was a lecturer in the Jin-Zhou Institute of Engineering. From 1989 to 1990 he was with the Automatic Control Laboratory of Gent University, Belgium. He is presently working towards a Ph.D. degree at Delft University of Technology, The Netherlands. His main research interests are robust and optimal control techniques.

Roumen L. Mishkov was born in V. Turnovo, Bulgaria, on 20 November 1954. He received the Dipl.Eng. degree in control engineering from the Higher Institute of Food Industry of Plovdiv, Bulgaria in 1980. He worked as a Development Engineer at the Central Institute for Complex Automation and at the Higher Institute of Food Industry during 1980–1982 and 1982–1986, resp-

ectively. Since 1986 he has been an assistant professor in the Control Systems Department, Technical University, Plovdiv, Bulgaria. His present research interests are in the areas of nonlinear systems, nonlinear observers, computer-aided control systems design.

Ivan Nagy received a degree in electrotechnical engineering from the Czech Technical University, Prague, in 1980 and the Ph.D. degree from the Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague, in 1983.

Since 1983 he has been a Senior Research Engineer in the same Institute. He is interested in adaptive control of random systems, especially in spline modelling of continuous systems and their discrete control with very short period of sampling.

Zigang Pan was born in Shanghai, People's Republic of China in 1968. He received his B.S. degree in automatic control from Shanghai Jiaotong University in 1990, and an M.S. degree in electrical engineering from University of Illinois at Urbana-Champaign in 1992. Currently, he is a research assistant in the Coordinated Science Laboratory of University of Illinois at Urbana-Champaign, pursuing a Ph.D. degree. His research interests are in H^∞-optimal control and control of singularly perturbed systems. He is the recipient of the 1992 Henry Ford II Scholarship Award at the University of Illinois.

P. N. Paraskevopoulos received the B.S. and M.S. degrees from the Illinois Institute of Technology in 1964 and 1965, respectively, and the Ph.D. degree from the University of Patras, Greece, in 1975, in Electrical Engineering. From 1969 to 1977 he was a research and development engineer at the NRC "Democritus", Department of Computer Science, Athens, Greece. From 1977 to 1984 he was professor of Automatic Control Systems in the Democritos University of Thrace, Xanthi, Greece. Since 1984 he has been professor of Control Systems in the National Technical University of Athens. His research interests are mainly in identification, system theory, design of control systems, and multidimensional systems. In these and other related areas he has published over 140 papers. Dr Paraskevopoulos is a senior member of the IEEE and a member of the Technical Chamber of Greece.

Anna Maria Perdon was born in Vicenza, Italy on 18 September 1948. She received the Doctor degree in Mathematics in 1972 from the University of Padova, Padova, Italy. She held various teaching and research positions at the University of Padova from 1972 to 1984 and from 1988 to 1991 and at the University of Genova from 1984 to 1987. She is now Associate Professor of Numerical Analysis at the University of Ancona. She was Fulbright Scholar in 1980 and 1987 and received a NATO Senior Fellowship in 1987. She was visiting researcher at the Econometric Institute of the Erasmus University of Rotterdam, The Netherlands, in 1979, Visiting Professor at the Ohio State University of Columbus, OH, in 1987 and Visiting Researcher at the Laboratoire Automatique of the CNRS of Nantes, France, in 1988. Her research interests include algebraic and geometric methods in systems and control theory.

Pedro L. D. Peres was born in Sorocaba, SP, Brazil, in 1960. He received the B.Sc. and M.Sc degrees in electrical engineering from the State University of Campinas, UNICAMP, in 1982 and 1985, respectively, and the "Doctorat en Automatique" degree from the University Paul Sabatier, Toulouse, France, in 1989. In 1990 he joined the Faculty of Electrical Engineering of the State University of Campinas, where he is currently a lecturer on control and optimization

---

Todor Ph. Proychev was born in Plovdiv, Bulgaria, on 5 November 1935. He received the Dipl. Eng. and Ph.D. degrees in electrical and control engineering from the Technical University of Sofia, Bulgaria in 1968 and 1977, respectively. From 1968 to 1974 he was an assistant professor of control engineering at the Control Systems Department of the Technical University in Sofia. During 1975–1980 he was an assistant professor, and in 1980–1985 associate professor, at the Department of Automation in the Higher Institute of Food Industry of Plovdiv. Since 1985 he has been an associate professor and Head of the Control Systems Department at the Technical University in Plovdiv. His present research interests are in the areas of optimal control, nonlinear systems and observers, and image processing.

Li Qiu received the B. Eng. degree in electrical engineering from Hunan University, Changsha, Hunan, China, in 1981, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, Ontario, Canada, in 1987 and 1990, respectively.

From 1982 to 1984, he worked as a full time teaching assistant at Hunan University. Since 1990, he

has held research positions at the University of Toronto, the Canadian Space Agency, and the University of Waterloo. At present, he is a research associate in the Institute for Mathematics and its Applications, University of Minnesota. His current research interests include linear and nonlinear control theory, robust control, digital control and the control of flexible structures.

Asok Ray earned the Ph.D. degree in Mechanical Engineering from Northeastern University, Boston, MA in 1976, and also graduate degrees in each of Electrical Engineering, Computer Science, and Mathematics. Dr Ray joined the Pennsylvania State University in July 1985, and is currently a Professor of Mechanical Engineering. Prior to joining Penn State, Dr Ray held research and academic positions at Massachussetts Institute of Technology and Carnegie-Mellon University as well as research and management positions at GTE Strategic Systems Division, Charles Stark Draper Laboratory, and MITRE Corporation. Dr Ray's research experience and interests include control and optimization of continuously varying and discrete-event dynamic systems in both deterministic and stochastic settings, intelligent instrumentation for real-time distributed processes, and design of fault-accommodating and robust control systems as applied to aeronautics and astronautics, power and processing plants, and autonomous manufacturing. Dr Ray has authored or co-authored over 180 research publications (including about 80 articles in referred journals) and is an Associate Fellow of AIAA, a Senior Member of IEEE and a member of ASME. Dr Ray is currently serving as an associate editor for two journals, namely, *Journal of Dynamic Systems, Measurement, and Control* and *International Journal of Flexible Manufacturing Systems*. Dr Ray is registered as a professional electrical engineer in the Commonwealth of Massachusetts.

Mario A. Rotea was born in Rosario, Argentina, on 6 August 1958. He received the degree of Electronic Engineer from the National University of Rosario, Argentina in 1983. He obtained the M.S. degree in Electrical Engineering and the Ph.D. degree in Control Science and Dynamical Systems from the University of Minnesota in 1988 and 1990, respectively. In 1989 he was awarded a Doctoral Dissertation Fellowship by the Graduate School at the University of Minnesota.

From 1983 to 1984, he was an Assistant Engineer at the Military Ammunition Factory "Fray Luis Beltran", Argentina. From 1984 to 1986 he was a Research Associate at the Institute of Technological Development for the Chemical Industry, Santa Fe, Argentina. In August 1990, Dr Rotea joined Purdue University, West Lafayette, where he is an Assistant Professor in the School of Aeronautics and Astronautics. His current research interests include robust multivariable control, optimal control, and applications of control theory to aerospace problems.

Prof. Rousan was born in Jordan on 4 September 1960. He received the B.S.E. in electrical engineering from Yarmouk University, Jordan, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from The Wichita State University, Wichita, KS, U.S.A., in 1987 and 1990, respectively.

He is currently an Assistant Professor of electrical engineering at Mutah University, Jordan. His research interest is in optimal control systems particularly pole placement and robustness.

Irwin W. Sandberg was born in Brooklyn, NY, on 23 January 1934. He received the B.E.E., M.E.E., and D.E.E. degrees from the Polytechnic Institute of Brooklyn in 1955, 1956, and 1958, respectively (Westinghouse Fellow 1956, Bell Telephone Laboratories Fellow 1957–1958).

He is presently a Professor of Electrical and Computer Engineering at The University of Texas at Austin, where he holds the Cockrell Family Regents Chair in Engineering, No. 1. Between 1958 and 1986 he was with Bell Laboratories, Murray Hill, New Jersey (as a Member of Technical Staff in the Mathematics and Statistics Research Center, and as Head of the Systems Theory Research Department from 1967 to 1972).

He has been concerned with analysis of radar systems for military defense, with synthesis and analysis of linear networks, with several studies of properties of nonlinear systems, and with some problems in communication theory and numerical analysis. His more recent interests include studies of the steady-state error performance of nonlinear control systems, and studies of the approximation and signal-processing capabilities of nonlinear networks.

He received the first Technical Achievement Award of the IEEE Circuits and Systems Society, is a Fellow of the IEEE, a fellow of the American Association for the Advancement of Science, an IEEE Centennial Medalist, a former Vice Chairman of the IEEE Group on Circuit Theory, and a former Guest Editor of the *IEEE Transactions on circuit theory* Special Issue on Active and Digital Networks. He has published extensively and holds nine patents.

Dr Sandberg is listed in several reference volumes, such as *Who's Who in America,* and is presently an advisor to *American Men and Women of Science*. He has received outstanding paper awards, an ISI Press Classic Paper Citation, a Bell Laboratories Distinguished Staff Award, and is a member of SIAM, Eta Kappa Nu, Sigma Xi, Tau Beta Pi, and the National Academy of Engineering of the United States.

George N. Saridis was born in Athens, Greece. He received the Diploma in Mechanical and Electrical Engineering from the National Technical University of Athens Greece in 1955, the MSEE and Ph.D. degrees from Purdue University, West Lafayette, IN, U.S.A., in 1962 and 1965, respectively. In 1988, he was certified as Manufacturing Engineer for Machine Vision by the Society of Manufacturing Engineers.

From 1955 to 1963 he was an instructor in the Department of Mechanical and Electrical Engineering of the National Technical University of Athens, Greece. From 1963 to 1981, he was with the School of Electrical Engineering of Purdue University. He was an instructor until 1965, Assistant Professor until 1970, Associate Professor until 1975, and Professor of Electrical Engineering until 1981. Since September 1981, he has been Professor of the Electrical, Computer and Systems Engineering Department and Director of the Robotics and Automation Laboratories at the Rensselaer Polytechnic Institute, in Troy, NY. In 1973 he served as Program Director of System Theory and Applications of the Engineering Division of the National Science Foundation, Washington DC. Since June 1988, he has been director of the NASA Center for Intelligent Robotic Systems for space exploration at RPI.

Dr Saridis is a Fellow of IEEE and a member of Sigma Xi, Eta Kappa Nu, the New York Academy of Science, the American Society of Mechanical Engineers, the Society of Photo-Optical Engineers, the American Society of Engineering Education, the American Society for the Advancement of Science, the American Association of University Professors and Amnesty International. He is also senior member of the Robotics International and Charter member of Machine Vision of the Society of Manufacturing Engineers. In 1972–1973 he served as the Associate Editor and chairman of the Technical Committee on Adaptive and Learning Systems and Pattern Recognition of the Control Systems Society of IEEE, Chairman of the 11th Symposium of Adaptive Processes, IEEE delegate to the 1973 and 1976 JACC, and Program Chairman of the 1977 JACC. In 1973 and 1979 he was elected member of the ADCOM and in 1986 he was appointed member of the Board of Governors of the IEEE Control Systems Society. In 1979–1981 he was appointed chairman of the Education Committee, and in 1986–1989 chairman of the Committee on Intelligent Controls of the same society. He was the International Program Committee chairman of the 1982 IFAC Symposium on Identification and Parameter System Estimation, in Washington DC, and the 1985 IFAC Symposium on Robotic Control in Barcelona, Spain. In 1974 and 1981 he was appointed Vice-Chairman of the IFAC International Committee on Education and in 1981–1984 the Survey Paper Editor of *Automatica*, the IFAC journal.

In 1983–1984 he was the Founding President of the IEEE Council of Robotics and Automation, and was elected member of the ADCOM of the IEEE Robotics and Automation Society in 1989 and 1990. He is also chairman of the Awards Committee of the same Society. In 1989 he served as member of the Panel on Intelligent Manufacturing of the National Research Council. In 1988 he was the General Co-chairman of the International Workshop on Intelligent Robots and Systems 'IROS '88' in Tokyo, Japan, Honorary Chairman of IPC of the 9th IFAC/IFORS Symposium on Identification, Budapest Hungary 1991, and Organizing Committee Chairman of the 'IROS'92', Raleigh NC, 1992.

Dr Saridis is the Editor of the series *Annuals on Advances in Robotics and Automation* of JAI Publications. He is the author of the book *Self-Organizing Control of Stochastic Systems*, co-author of the book *Intelligent Robotic Systems*, editor of the books *Advances in Automation and Robotics*, 1985, Vol. 1, 1990, Vol. 2 and co-editor of the books, *Fuzzy and Decision Processes*, *Proceedings of the 6th Symposium of Identification and System Parameter Estimation*, *Proceedings of the* 1985 *SYROCO and Knowledge Based Robotic Control*. He has written over 350 book chapters, journal articles, conference papers and technical reports. He has also presented more than 100 invited lectures.

Dr Saridis is the recipient of the IEEE Centennial Medal Award in 1984, and the IEEE Control Systems Society's Distinguished Member Award in 1989.

Michael Šebek was born in Prague, Czechoslovakia, in 1954. He received the Ing. degree in electrical engineering from the Czech Technical University, Prague, in 1978 and the CSc. (Ph.D.) degree in control theory from the Czechoslovak Academy of Sciences in 1981. Since 1981, he has been with the Institute of Information Theory and Automation, Prague, where he is currently Chief Scientist. He held visiting positions at the University of Padova, Italy, the University of Strathclyde, Scotland, the University of Toronto, Canada, and the University of Twente, The Netherlands. His current research interests are in robust control theory, $n-D$ systems and numerical algorithms for control. Dr Šebek is a Senior Member of IEEE.

Jenny Shen was born on 5 March 1961, in Beijing, People's Republic of China. She received a Bachelor of Science degree in Engineering Mechanics from Tsinghua University, Beijing in May 1984. After graduation, she studied and worked at the Institute of Applied Mathematics and Mechanics, Shanghai, People's Republic of China, until 1986. In May 1988 she received a Master of Science degree in Mechanical Engineering at San Diego State University, California. She went to Penn State in August 1988 and has studied in the Department of Mechanical Engineering for her Ph.D. degree. Her current research activities include robust design of multivariable systems and delay compensation of integrated communication and control systems.

David D. Sworder was born in Dinuba, CA. Upon earning his B.S. and M.S. degrees in electrical engineering from the University of California, Berkeley, he attended the University of California at Los Angeles where he completed the Ph.D. degree in engineering.

He was associated with the Department of Electrical Engineering at the University of Southern California from 1964 to 1977. Since 1977, he has been Professor of Engineering first in the Department of Applied Mechanics and Engineering Science and more recently, the Department of Electrical and Computer engineering at the University of California, San Diego. He is also Associate Dean of Graduate Studies and Research for the campus. Dr Sworder has interest in guidance and control of systems subject to stochastic influences. He has published widely in this area and has worked closely with both industrial and governmental laboratories on the synthesis of tracking algorithms containing unusual sensor configurations.

Mario Sznaier received the Ingeniero Electronico and Ingeniero en Sistemas de Computacion degrees from the Universidad de la Republica, Uruguay in 1983 and 1984, respectively and the MSEE and Ph.D. degrees from the University of Washington in 1986 and 1989, respectively. He spent 1990 as a Research Fellow in Electrical Engineering at the California Institute of Technology. In 1991 he joined the Department of Electrical Engineering at the University of Central Florida where he is currently an Assistant Professor. His research interest include Optimal Control Theory, Robust Control, applications of computers to control systems, and stability and robustness issues in Intelligent Control.

Yu Tang was born in Beijing, China, in 1960. He received B.Eng. in Computer Engineering, M.Eng. and Ph.D., both in Electrical Engineering in 1984, 1985 and 1988, respectively, all from the National University of Mexico, Mexico. In 1989, he joined the Division de Estudios de Posgrado, Facultad de Ingenieria at the National University of Mexico, where he is now a Professor. Dr Tang has also been a part-time professor at the Centro de Investigaciones y Estudios Avanzados, Mexico, since 1990. His current research interests include adaptive control, robust control, system identification, signal processing and robotics.

Gang Tao received the BSEE from University of Science and Technology of China, MSEE, MSCE, MSAMATH and Ph D in EE from University of Southern California, in 1982, 1984, 1987, 1989, respectively.

From August 1989 to May 1991 he was a visiting assistant professor in Electrical and Computer Engineering Department of Washington State University Since June 1991 he has been an assistant research engineer in Department of Electrical and Computer Engineering of University of California at Santa Barbara.

His research and teaching interests are mainly in adaptive control theory and applications, linear and nonlinear control systems, robotics, applied mathematics.

Arja Toola received the M.Sc. degree in engineering from the Tampere University of Technology in 1985, and the Ph.D. in 1992. She spent three years with Imatra Power Company, working on problems of reliability of automation. For the past five years she has been with Technical Research Centre of Finland, Safety Engineering Laboratory, working on safety of computer controlled processes.

Herbert Tulleken was born in The Hague, The Netherlands on 7 April 1956. In 1977 he received the equivalent of an B.Sc. degree in agriculture. After a one-year assignment at an agricultural research institute, he switched over to the field of applied mathematics.

In 1984 he received the equivalent of an M.Sc. degree (*cum laude*) at Twente University of Technology, with specialization in Stochastic System and Control Theory. Since then he has been employed by Koninklijke/Shell Laboratorium, Amsterdam (Shell Research B.V.) in the department of Mathematics and System Engineering (sections Process Control and, from 1990, Operational Analysis and Optimization). His current research efforts are in Markov modelling and Statistical Quality Control; his interests range from identification and control to stochastic operational planning and scheduling problems.

A Doctoral thesis in the field of model identification has been completed and defended in 1992 at Delft University of Technology.

Lilian Xu is a Ph.D. candidate at the University of Texas at Austin. She received B.S. degrees in both Physics and Electrical Engineering from the University of Missouri, and an M.S degree in Electrical Engineering from the University of Texas at Austin. Her current research interests are in the area of nonlinear systems. She is a member of Eta Kappa Nu and Tau Beta Pi

---

Oden Yaniv was born in Israel in 1950. He recieved the B.Sc. degree (Mathematics and Physics) from the Hebrew University, Jerusalem, in 1974, and M.Sc. (physics) and Ph.D. (applied mathematics) degrees from the Weizmann Institue of Science, Rehovot, Israel in 1976 and 1984, respectively. His industrial experience includes being a development engineer at the Israel Aircraft Industries 1979–1980, and serve as a Senior Control Engineer at Tadiran (System Division), Holon, Israel 1983–1987. Since 1987 has has been at the Tel-Aviv University, Israel. His main research interests include synthesis of uncertain multi-input–multi-output feedback systems, linear as well as nonlinear.

Hao Ying was born in Shanghai, China. He received B.S. and M.S. degrees in Electrical Engineering in 1982 and 1984 from China Textile University, Shanghai. He received the Ph.D. degree in Biomedical Engineering in 1990 from University of Alabama at Birmingham, U.S.A. While studying for the Ph.D. degree, he worked as a Research Fellow in the Kemp–Carraway Heart Institute and as an Instructor in the Department of Mathematics at UAB

He is presently an Assistant Professor in the Department of Physiology and Biophysics, University of Texas Medical Branch. His research interests include theory and application of fuzzy control, and application of expert systems and neural networks. The focus of his current research is on intelligent control in biomedical environments.

Zhihong Zhang was born in Anhui, China, in 1961. He received the B.S. degree from the China University of Science and Technology, Hefei, China, in 1983, and the M.S.E. and Ph.D. degrees from the University of Michigan, Ann Arbor, in 1986 and 1990, respectively, all in electrical engineering.

Since 1990, he has been with the General Motors Systems Engineering Center, Troy, MI, where he is presently a Senior Research Engineer in the Department of Systems Analysis and Information Management. His current interests include automotive applications of robust control theory, neural networks and fuzzy systems, and decision theory.

Dr Zhang is a member of IEEE Control Systems Society, Tau Beta Pi, and an associate member of SAE.

Automatica, Vol. 29, No. 2, p. 585, 1993
Pergamon Press Ltd.

Addendum to List of Reviewers for *Automatica*

The following names were omitted from the List of Reviewers for *Automatica*, 1992 (Vol. 28, No. 6, 1992) as they were unavailable at the time of going to press.

1992
Andeen, G.
Ang, Marcelo
Annaswamy, A.M.
Atkeson, C.G.

Bagchi,
Bai, E.W.
Barton, G.
Bernard, C.
Besharah-rad A.
Boje, E.
Bokor, J.
Brdys, M.
Brogliato, B.B.
Bryant, G.

Canadus-de-Witt, C.
Carlsson, B.
Chen, Y.C.
Cheng, Shao
Cooper, D.J.
Cotsaftis, M.

Deistler, M.
De Mathelin, M.
De Moor, B.
De Vries, D.K.
DiMascalo, N.
Dogan, M.
Dumont, G.A.

El Shafei, A.L.

Feng, Weiming
Fleming, P.J.
Fredrick, D.K.
Fukuda, T.

Gersch, W.
Graebe, S.
Guidorzi, R.P.
Gustaffson, F.

Hakvoort, R.G.
Hang,
Harvey, C.A.
Heemink, A.W.
Heij, C.
Helmersson, A.
Hjalmarsson, H.
Hulkó, G.

Jacobs, O.L.R.
Jing, Sun

Kaman, E.W.
Karlheinz, P.
Kennedy, R.
Keuchel, U.
Keyser De, R.M.C.
Khuen, H.W.
Krishnamurpy, V.
Kunisch, K.
Kuznetsov, A.
Kwakemaak, H.

La Marie, R.O.
Lau, M.
Lawrence, D.A.
Lee, C.C.
Lim, K W.
Linkens, D.A.
Lototsky, V.A.

MacCluer, C.R.
Mahony, R.
Mäkilä, P.
McGinnie, B.P.
Mendel, J.M.
Morris, A.J.
Motadi, C.

Nakamizo, T.
Niederlinski, A.
Ninness, B.M.
Nordström, K.

Ober, R.
Ohmori, H.
Olsson, G.

Rad, Ahmad
Ramadorai, A.
Rao, G.P.
Rawlings, J.B.
Rey, G.J.
Roberts, P.D.

Salgado, M.
Sciacca, L.
Seraji, H.
Soeterboek, A.R.M.
Soh, C.B.
Solo, V.
Sundararajan, N.
Sunuhara,

Tan, S.H.
Tariq, M.
Tay, T.T.
Teo, Y.T.
Teoh, E.K.
Thibault, J.
Timmerman, Marc
Tulleken, H.J.A.F.
Tzafestas, S.

Ungar, L.H.
Uosaki, K.

Vajta, M.
Van Breusegem, V.
Van den Hoef, P.
Van Overschee, P.
Van Shuppen, J.H.

Wang, Qing-Guo
Wee, Liang-Boon
Wen, Changyun
Wiberg, P.M.

Xu, Jian-Xin

Yasuda, T.
Young, P.C.

Zaid, F. Ahmed
Zang, C.S.
Zhao, X-H.-
Zhu, Y.C.
Zimmermann, M.